Structural studies on β-lactoglobulin and cyclophilin 3

Su-Ying Wu

PhD Thesis
University of Edinburgh
1999

I declare to have written this thesis myself and to have personally done the work described.

Acknowledgements

I would firstly like to thank both of my supervisors, Dr. Lindsay Sawyer and Prof. Malcolm Walkinshaw, for all the guidance and encouragement throughout the course of my studies.

I am deeply indebted to Dr. Sharon Brownlow, Dr. Dimitri Alexeev and Dr. Paul Taylor for the essential scientific teaching and invaluable suggestions.

I would also like to thank Dr. Carl Holt and Elaine Little in Hannah Research Institute for their technical assistance and friendship.

Thanks also go to the all members in the Structural Biochemistry Group, especially, Jacqueline and George, for making my studies enjoyable.

I am particularly indebted to all the friends, especially Ting-Huei, and the Taiwanese fellowship group for their precious friendship during my time in Edinburgh.

Finally, special thanks to my parents and my brother for their love and support from the very beginning and always.

Abstract

X-ray crystallography has played an essential role in the field of molecular biology and biochemistry as the protein structures obtained by this method provide a variety of information with which to examine the interactions between proteins and ligands and consequently to help understand the biological functions of the proteins. In this thesis, the structures and protein-ligand binding interactions of two proteins, β-lactoglobulin (BLG) and cyclophilin 3 (CYP3), are studied by x-ray crystallography.

BLG is a small globular protein, belonging to the lipocalin family. It is a commercially important whey protein of nutritional value. However, the biological function is not yet clear. The structures of BLG in three space groups have been solved. However, the refinement of the lattice Y structure has shown problematic. In the current study, a new data set has been collected at low temperature (100K) and the structure has been refined to 2.0Å with the R-factor of 25.7% (R_{free}=31.0%). The R-factor remains around 26% although many methods have been tried to lower it. The possible reasons that hinder the successful refinement are discussed.

In common with many of the other members of the lipocalin family, BLG binds a variety of hydrophobic ligands. It appears possible that there are two distinct binding sites per monomer. By comparison with other members of the family, there is a probable binding site in the central calyx of the protein formed by the eight antiparallel β-strands. In this thesis, cocrystallization of BLG with palmitate has been successful and the structure of palmitate-bound BLG has been solved and refined at a resolution of 2.5Å to an R-factor of 20.4% (R_{free} = 24%). The clear electron density map shows the ligand binds inside the calyx and lies parallel to β-strand E. The palmitate adopts an almost fully extended conformation with the hydrophobic tail stretching into the bottom of the calyx. The carboxyl group of palmitate forms two hydrogen binds with Lys60 and Lys69 at the entrance of the calyx. This the first direct observation of a ligand binding to BLG.

Cyclophilins are known as immunophilins as they can bind tightly to cyclosporin (CsA), which is a immunosuppressant drug used to prevent organ rejection after transplant operations. The complex, cyclophilin-CsA, can inhibit the activation of the immune cells. In addition, cyclophilins also possess the peptidyl-prolyl *cis-trans* isomerase (PPIase) activity, which can catalyse the *cis→trans* isomerization of proline-containing peptide bonds and consequently accelerate the refolding of some proteins in the cell. Recently, eleven cyclophilin isoforms have been cloned and sequenced from the free-living nematode *Caenorhabditis elegans* (*C.elegans*). Cyclophilin 3 (CYP3) is one of the most abundant expressed isoforms in this family. The structure of CYP3 has been refined to 1.8Å with R-factor of 20.8% (R_{free} = 28.4%). The structure is similar to the well-characterised human cyclophilin A. It contains a eight-stranded antiparallel β-barrel, similar to, but distinct from, BLG, capped with two α-helices. However, CYP3 has an additional loop composed of seven residues (48KSGKPLH54), which is absent in CYPA. This loop is held tightly by a specific network of hydrogen bonds from the side chain of Glu83. The biological role of this additional loop is still unclear although it may provide a distinctive recognition feature of the *C.elegans*-cyclophilin family as this inserted loop is particularly common in this family.

X-ray data for 120mM Ala-Pro bound to CYP3 has been collected and refined at 1.9Å resolution to an R-factor of 18.3% (R_{free} = 24.4%). The clear density map shows Ala-Pro binds in a groove on the surface of the protein. The side chain of proline makes strong hydrophobic contacts with the hydrophobic pocket of CYP3. In addition, there are seven strong hydrogen bonds involved in the ligand-protein interactions.

Finally, the structure of CYP3 complexed with Ala-Pro has been used as a model to study the ligand binding behaviour *"in crystallo"*. Data from a series of native crystals soaked in solutions with different concentrations of Ala-Pro have been collected. The complexed structures with different ligand concentrations have been refined and the occupancy of Ala-Pro in each structure has been determined. From the occupancy of Ala-Pro, the dissociation constant of Ala-Pro binding to CYP3 has been calculated. The dissociation constant obtained by x-ray crystallography is 27.7mM, comparable with the result obtained from solution studies.

Abbreviations

acAAPAamc	acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin
$Arg62_{(L)}$	Arg62 conformation with Ala-Pro binding to CYP-3
$Arg62_{(N)}$	Arg62 conformation without Ala-Pro binding to CYP-3
BLG	β-lactoglobulin
BLGX	the structure of BLG in the lattice X form
BLGY	the structure of BLG in the lattice Y form
BLGZ	the structure of BLG in the lattice Z form
C. *elegans*	*Caenorhabditis elegans*
CA_{151}	the N-terminal domain (residues 1 to 151) of the HIV-1 capsid protein
CsA	cyclosporin
CYP3	cyclophilin 3
DMSO	dimethylsufoxide
FABP	fatty acid binding protein
hCYPA	human cyclophilin A
K_{dc}	dissociation constant *"in crystallo"*
PPIase	peptidyl-prolyl *cis-trans* isomerase
PR	progesterone receptor
RBP	retinol binding protein
rms	root mean square
sAAPFna	succinyl-Ala-Ala-Pro-Phe-p-nitroanilide

Table of contents

List of figures

List of tables

1. Introduction

This chapter describes the biological characteristics of β-lactoglobulin (BLG) and cyclophilin 3 (CYP3), whose structures and ligand-protein binding interaction are studied by x-ray crystallography in this thesis. The introduction to BLG will be given in section 1.1 followed by the introduction to CYP3 in section 1.2. Finally, the aim of this thesis will be described in section 1.3.

1.1 β-Lactoglobulin (BLG)

1.1.1 Introduction

β-Lactoglobulin (BLG) is the major whey protein found in the milk of many mammals and is expressed in the mammary gland during late pregnancy and lactation (Larson, 1979). It is a small soluble protein and is remarkably acid stable. It remains intact even at pH 2. The structure of the protein when crystallised at low pH is similar to that at pH 7, with the exception of some loops which are disordered at low pH (Ragona *et al.*, 1997; Fogolari *et al.*, 1998).

BLG has been detected in the milk of ruminants, for example, cows and sheep, and also in non-ruminants, for example, pigs and dogs. However, it is absent from the milk of humans, rabbits and mice. Since its isolation in 1934 (Palmer, 1934) and mainly because of its abundance and ease of purification, bovine BLG has been extensively studied using a variety of techniques and a lot of physico-chemical properties have been investigated (Tilley, 1960; Townend *et al.*, 1969; McKenzie, 1971; Green *et al.* 1979; Hambling et al., 1992).

In general, at room temperature and physiological pH, BLG exists as a dimer in the milk of ruminants while it exists as a monomer in the milk of non-ruminant species. However, ruminant BLG dissociates into monomers at pH values below 3 and above 8. Each monomer consists of 162 amino acids and has one free cysteine and two disulphide bridges (Grosclaude *et al.*, 1976). Generic variants appear to exist in all species. In cows, the predominant forms are the A and B variants. The structure of BLG contains a β-barrel formed by eight antiparallel β-strands, labelled as strand A,

B, C, D, E, F, G, H. A 3 turn α-helix is on the outer surface of the β-barrel and is followed by the ninth β-strand, labelled as β-strand I.

BLG is a member of the lipocalin family (Flower, 1996). Many members of this family are able to bind small hydrophobic molecules and BLG is no exception. It is well known a variety of ligands can bind to BLG, for example, retinol (Fugate and Song, 1980), fatty acid (Spector and Fletcher, 1970; Frapin *et al.*, 1993), alkanes, aromatic compounds and some biological molecules, such as heme-CO (Marden *et al.*, 1994). However, the site of ligand binding and the mechanism of binding to BLG have remained mysterious for many years. Other lipocalin members, such as retinol-binding protein (Newcomer *et al*, 1984; Cowan *et al*, 1990), bilin binding protein (Huber *et al*, 1987), odorant binding protein (Monaco *et al.*, 1990) and fatty acid binding protein (Hodsdon and Cistola, 1997a, b), bind ligands inside the calyx formed from eight (or ten) antiparallel β-strands.

The biological function of BLG is unknown. The amino acid composition and quantities present in ruminant species provide evidence to support a nutritional role. However, the ligand binding properties that have emerged, coupled with the structural similarity of BLG to retinol-binding protein and possibly even FABP, have led to suggestions of its having a role in the transport of ligands like retinol or fatty acids (Newcomer *et al.* 1984). The majority of the ligands that have been examined are hydrophobic so that a transport role, in keeping with other lipocalins, is certainly possible although no real proof of this has been published to date.

Moreover, it has been reported that bovine BLG increases the activity of ruminant pregastric lipase (Perez *et al.* 1991). Since this enzyme is involved in the digestion of milk lipid, it is assumed BLG may play an essential role in milk fat digestion in the new-born animal by increasing the activity of pregastric lipase.

1.1.2 Distribution

The composition of milk varies from species to species. There are also seasonal variations while the composition of milk changes in different seasons, presumably implying the dairy habit of the animal. Some sensitive methods, such as the

immunochemical cross-reaction between BLG and bovine BLG antiserum, have been developed to detect the presence or absence of BLG in the milk of many species. BLG has also been found in many ruminant species, such as, cow (Palmer, 1934), sheep (Bell and Mckenzie, 1967), goat (Townend and Basch, 1968), water buffalo (Lyster, 1972) and red and Virginian white-tailed deer (McDougall and Stewart, 1976). Not only in the milk of ruminant animals, BLG is also found in non-ruminant animals, such as horse (Bell *et al*, 1981), red and eastern kangaroos (McKenzie *et al*, 1983), dolphins, manatees (Pervaiz and Brew, 1986) and cats (Halliday *et al.*, 1993). However, it is of interest to note that BLG in human milk is not detectable.
Alignment of the published sequences of BLG shows there is a very high degree of homology throughout ruminant BLG. The sequence identity among ruminant BLG is greater than 93%. However, the sequence identity is much lower among non-ruminant BLG (Kolde, 1983). Some non-ruminant species, such as the pig I and horse II BLG, shares only 37.7% identity with one another.

1.1.3 Genetic variants

Bovine BLG reveals heterogeneity and nine variants have been found so far. The comparison of the amino acid sequence of different genetic variants of bovine BLG is shown in Table 1-A.
BLG-A and BLG-B are the most abundant (Aschaffenburg and Drewry, 1955). These two variants have two residue differences, residue 64 and residue 118. As shown in Table 1-A, residues 64 of BLG-A and BLG-B are Asp and Gly, respectively while residue 118 of BLG-A and BLG-B are Val and Ala, respectively. Therefore, variant A is more negative than B so that they could be separated under the influence on the applied electric field. BLG-C was identified in the milk of Jersey cows (Bell and Mckenzie, 1967) and South Africa Nguni cattle (Osterhoff and Pretorious, 1966). BLG-C reveals only one residue difference when compared to BLG-B. The residue 59 is Gln and His in BLG-B and BLG-C, respectively. BLG-D (Brignon *et al.*, 1969) is found in Danish Jersey cows and replaced by a single residue at position 45, Glu to Gln, with respective to BLG-B.

Table 1-A: Comparison of the amino acid sequence of different genetic variants in bovine BLG

Res-idue No.	BLG-A	BLG-B	BLG-C	BLG-D	BLG-Dr	BLG-E	BLG-F	BLG-G	BLG-W
1		V							
11		D							
20		Y							
28		D			N*				
45		E		Q					
50		P					S		
53		D							
56		I							L
59		Q	H						
64	D	G			D				
78		I						M	
118	V	A			V				
129		D					D/Y+		
130		D					D/Y+		
148		R							
150		S							
158		E				G	G	G	
162		I							

Key:

The blank represent the same residue at the same position compared to BLG-B

D/Y+ : it is uncertain which residue is at this position

N*: represents an Asn residue with a carbohydrate moiety attached. It is within the sequence -NIS- which is typical of a N-glycosylation site

Three genetic variants E, F, G (Bell *et al.*, 1981) were isolated from Bali cattle, *Bos javanicus*. It has been suggested that the E variant appears to have the same sequence as the BLG from yak milk, as position 11 in the latter is now thought to be Asp (Bell *et al.*, 1981) and not Asn (Grosclaude *et al.*, 1976). BLG-E contains a single substitution compared to BLG-B. The residue 158 is Gly in BLG-E instead of Glu. BLG-F has one residue difference compared to BLG-E. Residue 50 is Ser and Pro in BLG-F and BLG-E, respectively. However, there are two unidentified residues, residue 129 and 130, in BLG-F. BLG-G also has one residue difference compared to BLG-E. Residue 78 is Met and Ile in BLG-G and BLG-E, respectively. BLG-W (Godovac-Zimmermann *et al.*, 1990; Krause *et al.*, 1988) was isolated from a rare breed of Bavarian and Murnau werdenfelder cattle. It has an electrophoretic mobility close to that of the B variant but differs by a single Ile to Leu substitution at position 56. Cross breeds of zebu and Australian Droughtmaster, were shown to produce BLG-A, BLG-B and a new variant BLG-Dr. BLG-Dr is the only BLG known to be N-glycosylated (Bell *et al.*, 1970). These genetic variants are different in their isoelectric points and can be separated using gel electrophoresis.
Other ruminant BLGs also exhibit genetic polymorphism. Red deer possess two variants (McDougall and Stewart, 1976) while sheep have variants A, B (Bell and McKenzie, 1967) and C (Erhardt *et al.*, 1989).

1.1.4 Native structure

1.1.4.1 Crystal forms

A variety of crystal forms of BLG have been produced from a variety of different conditions including pH, salt concentration and presence of ligand (Crowfoot and Riley, 1938; Senti and Warner, 1948; Green *et al.*, 1956; Aschaffenburg *et al.*, 1965; Bolognesi *et al.*, 1979). In fact, there have been 18 crystal forms identified to date. The majority of studies have concentrated on three salted-out crystal forms: lattice X (Brownlow *et al.*, 1997), lattice Y (Papiz *et al.*, 1986) and lattice Z (Monaco *et al.*, 1987) as they are amenable to crystallisation at physiological pH. These crystal forms were crystallised on either sides of the Tanford transition (Tanford *et al.*, 1959),

which occurs between pH 6 to 8. The details of the Tanford transition will be discussed in section 1.1.7.

The Tanford transition is also known as the N → R transition where N and R represent the protein conformation at pH 6-6.5 and the conformation at pH 7.8, respectively. Lattice X was crystallised at pH 6.5, corresponding to the N conformation in solution while lattice Y was crystallised at pH 7.5, corresponding to the R conformation in solution. Lattice Z was usually crystallised at a pH above 7 and sometimes this crystal form is found in the same drop with lattice Y. However, it has been recently reported, lattice Z can be crystallised at a number of different pHs including pH 6.2, pH 7.1 and pH 8.2 (Qin *et al.*, 1998a). The space groups and unit cell dimensions of these three crystal forms are summarised in Table 1-B.

Table 1-B. The space groups and unit cells of three crystal forms of BLG

	Space group	Unit cell	pH	Reference
lattice X	P1	a=37.8Å b=49.6Å c=56.6Å α=123.4° β=97.3° γ=103.7°	6.5	Brownlow *et al.*, 1997
lattice Y	$B22_12$	a=55.38Å b=66.72Å c=81.69Å α=β=γ=90.0°	7.6	Papiz, 1982 Papiz *et al.*, 1986
lattice Z	$P3_221$	a=b=53.75Å c=111.5Å α=β=90.0° γ=120.0°	6.2	Qin *et al.*, 1998a
lattice Z	$P3_221$	a=b=53.96Å c=112.41Å α=β=90.0° γ=120.0°	7.1	Qin *et al.*, 1998a
lattice Z	$P3_221$	a=b=54.29Å c=113.1Å α=β=90.0° γ=120.0°	8.2	Qin *et al.*, 1998a

1.1.4.2 Recent history of crystallography studies on BLG

BLG has been studied by x-ray crystallography for many years since in 1979 when the first BLG structures of X, Y, Z lattices were reported (Green *et al.*, 1979). In these 6Å structures, BLG was shown to exist as a dimer. However, the positions of the intramolecular disulphide bridges were unknown.

In 1986, a higher resolution structure was published by Edinburgh and Leeds groups (Papiz *et al.*, 1986). In this 2.8Å resolution structure, the main fold of BLG was recognised. 8 out of the 9 β-strands form a β-barrel and the 3 turn α-helix is outside of the β-barrel. This special β-barrel motif is very similar to the previously solved structure of retinol-binding protein (Newcomer *et al.*, 1984). The positions of disulphide bridges were easily recognised in this paper (Papiz *et al.*, 1986). Cys106 and Cys119 from strands G and H, respectively, form one disulphide bridge while Cys66 from strand D and Cys160 in the C-terminal form the other.

One year later, Monaco (Monaco *et al.*, 1987) built a new structure model of lattice Z at 2.5Å resolution by obtaining the phases calculated from the multiple isomorphous replacement (MIR). However, the model had poor geometry as reported by a structure quality check program PROCHECK and unsatisfactory crystal contacts. Therefore, the accuracy of this model is questionable.

In 1993, Joao Cabral re-collected the derivative data for lattice Z to 3Å and re-solved the structure by the MIR method (Cabral, 1993).

In 1997, Brownlow (Brownlow *et al.*, 1997) used the new structure of lattice Z solved by Cabral (Cabral, 1993) as the search model and solved the high resolution structure of lattice X by molecular replacement to 1.8Å. In this new model, the overall fold is the same as the old lattice Y model. However, there are many differences in the loops and there are four additional short α-helical turns that are absent in the old lattice Y model. The most obvious difference is in the AB loop, which extends 5 residues compared to the old model. The longer AB loop contributes greatly to the stabilisation of the interface between the two monomers.

The shift of 5 residues in AB loop affects the position of 3 β-strands (A, B, C) and consequently results in shorter CD loop. The first residue in D strand was replaced

back in register with the previous lattice Y model as a consequence of the shorter CD loop.

1.1.4.3 Overall structure

The crystal structure of BLG reveals to be an eight-stranded antiparallel β-barrel arranged to form a conical central calyx or cavity in which the hydrophobic ligand is expected to be located (Brownlow *et al.*, 1997; Qin *et al.*, 1998a). The overall structure of BLG is shown in Figure 1-1. The secondary structure and corresponding residues based on a monomer of BLG in lattice X form (BLGX) are listed in Table 1-C.

Table 1-C. The secondary structure and corresponding residues of a monomer of BLGX

Residues range	secondary structure	Residues range	secondary structure
Gln5-Asp11	N-terminal	Asn90-Thr97	**Strand F**
Ile12-Val15	***3₁₀ Helix***	Asp98-Lys101	FG loop
Ala16-Trp19	Loop	Tyr102-Glu108	**Strand G**
Tyr20-Ala26	**Strand A**	Asn109-Glu112	GH loop
Ser27-Asp28	AB loop	Pro113-Gln115	***3₁₀ Helix***
Ile29-Leu31	***3₁₀ Helix***	Ser116-Leu117	GH loop
Leu32-Val41	AB loop	Val118-Val123	**Strand H**
Tyr42-Pro48	**Strand B**	Arg124-Asp129	Loop
Thr49-Asp53	BC loop	Asp130-Lys141	***Helix***
Leu54-Trp61	**Strand C**	Ala142-His146	Loop
Glu62-Glu65	CD loop	Ile147-Ser150	**Strand I**
Cys66-Lys75	**Strand D**	Phe151-Asn152	Loop
Thr76-Ala80	DE loop	Pro153-Leu156	***3₁₀ Helix***
Val81-Ala86	**Strand E**	Glu157-Cys160	C-terminus
Leu87-Glu89	EF loop		

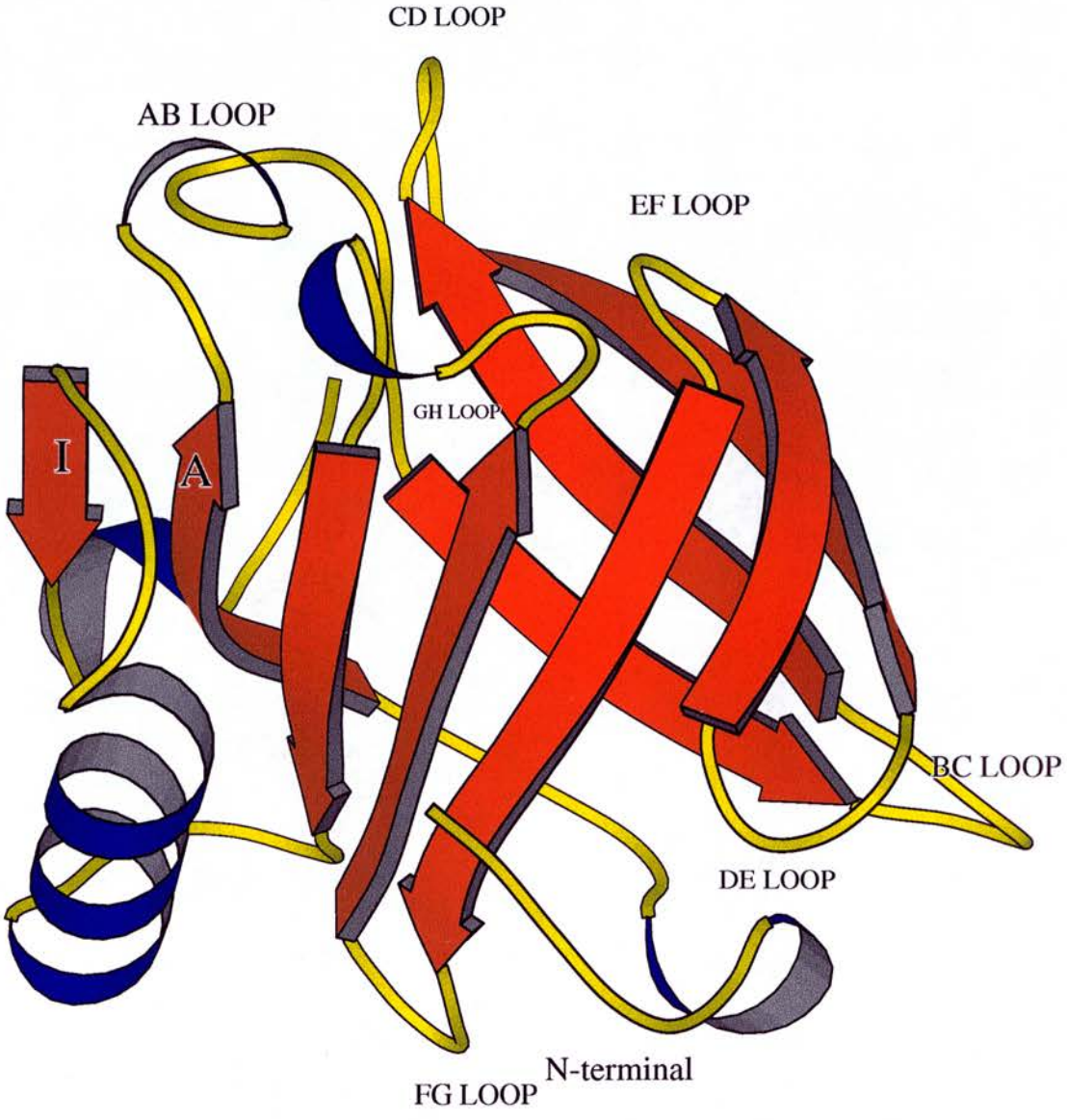


Figure 1-1.Diagram of a monomer of BLGX. The nine strands are shown in red, the helices are shown in blue and the joining loops are shown in yellow. Strand A, I and loops are labelled. This diagram was prepared by MOLSCRIPT (Kraulis, 1991)

The first four β-strands (A, B, C, D) form one side of the calyx while the other four β-strands (E, F, G, H) form the other. Strand A is bent through a right angle at Leu22 and contributes to both sides of the calyx. The 3-turn α-helix is located outside the β-barrel and is followed by β-strand I, which forms the dimer interface along with the AB loop. There are twelve H-bonds at the interface, four of them occur in β-strand I between two adjacent monomers. The other eight hydrogen bonds are formed in the AB loop between the two adjacent monomers (Brownlow *et al.*, 1997). In addition, there are four short 3_{10} α-helix turns which occur in the N-terminus, AB loop, GH loop and C-terminus.

There are five cysteine residues in BLG. Four of them form two disulphide bridges as mentioned in section 1.1.4.2. The free cysteine, Cys121, was found to be buried in the interface of β-sheet and α-helix, which is consistent with the conclusion of Townend (Townend *et al.*, 1969).

1.1.5 Studies on specific amino acids

The environments of some specific amino acid of BLG have been studied under various conditions using difference spectroscopy and chemical modification (Townend *et al.*, 1969). These specific amino acids include 4 tyrosines, 2 tryptophans and the free cysteine. Four tyrosines (Tyr20, Tyr42, Tyr99 and Tyr102) and two tryptophans (Trp19 and Trp61) will be discussed here. The free cysteine will be discussed in section 1.1.7.

(a) *Four tyrosines (Tyr20, Tyr42, Tyr99 and Tyr102)*:

BLG contains four Tyr residues per monomer and their accessibility has been studied by titration with N-acetylimidazole and cyanuric fluoride (Townend *et al.*, 1969). It is found that the two tyrosines are able to acetylate with N-acetylimidazole at pH 7.5 whereas one tyrosine is totally unavailable for acetylation. The fourth tyrosine is partially hindered and only becomes available after partial denaturation of BLG at pH 10.8. These four tyrosines can therefore be categorised as: two are exposed, one is partially buried and the rest one is completely buried.

This categorisation is consistent with the structural analysis of BLG (Qin *et al.*, 1998a). By calculating the solvent accessibility of the sidechains of two tyrosines, Tyr20 and Tyr99, are identified as the two exposed tyrosines with the solvent accessibility of 40Å^2 and 30Å^2, respectively. Tyr102 and Tyr42 are identified as the partially buried and buried tyrosine with the solvent accessibility of 10Å^2 and 0Å^2, respectively.

(b) *Two tryptophans (Trp19 and Trp61):*

There are two tryptophans in BLG, Trp19 and Trp61. Trp19, located at the bottom of the calyx, is inaccessible to solvent while Trp61, in the extremely mobile loop at the mouth of the calyx, is exposed.

Trp19 is the only conserved residue throughout the lipocalin family, implying this residue may play an important role shared by the lipocalin family. A study using site-directed mutagenesis reported Trp19 to be important in stabilising the protein structure (Katakura *et al.*, 1994). The concentration of guanidine hydrochloride (as a denaturant) for the loss of CD intensity at 220 nm was 2.5M for wild-type BLG. As for W19Y mutant, the denature concentration decreased to 1.2M, indicating W19Y mutant was destabilised by this substitution.

Moreover, Cho *et al* (1994) genetically modified Trp19 to Ala19 and found, upon binding of retinol, the association constant of the W19A mutant was decreased to 73% compared to the wild type. This result reveals the Trp19 has also an effect on ligand binding.

1.1.6 Binding studies by other methods

Since the late 1940's, many binding studies have been on BLG since 1946 when the first compound, sodium lauryl sulphate(SLS), was reported to bind to the BLG (Bull *et al.*, 1946). As mentioned in section 1.1.1, There are a variety of ligands can bind to BLG. In most cases, BLG binds 1 molecule of ligands per monomer with an association constant varying from 10^7 to $10^2(M^{-1})$. Among them, retinol and fatty acid have the strongest binding affinity and have been investigated extensively because their binding indicates a possible biological function for BLG. Therefore, the

binding of fatty acid and retinol to BLG will firstly be discussed, followed by some other compounds.

1.1.6.1 Fatty acid

In 1970, fatty acid was first reported to bind to BLG by Spector and Fletcher (1970) with an association constant around 10^5 (M^{-1}).

The strength of binding depends on the length of the carbon chain. Caprylic acid (C8:0) and capric acid (C10:0) do not bind (Frapin *et al.*, 1993). If C>12, the association constant is higher if the carbon chain is longer, which suggests the binding process involves hydrophobic interactions. For C>16, the strength of binding decreases as the carbon chain length increases. Conclusively, the strength order is: C16:0> C18:0>C14:0> C20:0 > C12:0. From this order, it seems optimum number of carbon atoms to fit into the binding pocket is 16. Furthermore, the binding affinity of fatty acid to BLG decreases if the carboxyl group of fatty acid has been removed or modified (Spector and Fletcher, 1970), indicating the polar interactions are also important in the binding process. Therefore, it was proposed the binding site for fatty acid contains a hydrophobic pocket with a positively charged amino acid at the entrance of the pocket (Spector and Fletcher, 1970; Perez *et al.*, 1989). This hypothesis has been proven by the structural studies described in this thesis. The complexed structure of palmitate binding to BLG reveals that palmitate binds inside the hydrophobic pocket (calyx) formed by the β-barrel. Lys60 and Lys69 at the entrance of the calyx are hydrogen bonded to the carboxyl group of palmitate. The detail of the complexed structure of palmitate binding to BLG will be discussed in chapter 3.

Comparing saturated and unsaturated fatty acid binding, the unsaturated fatty acids bind more weaker than the saturated ones. This indicates the structural constraints caused by C=C bonds weaken their binding to BLG (Frapin *et al.*, 1993).

Moreover, modifying BLG by alkylation and esterification also decreases the binding affinity as methylation of lysine residues in BLG and methyl esterified BLG does not bind palmitate at neutral pH (Frapin *et al.*, 1993). This result shows the partial change of BLG alters the fatty acid binding site.

It has been reported that the binding of fatty acids to BLG increases its resistance to proteolytic degradation and thermal denaturation (Puyol *et al.*, 1994) as the denaturation temperature of BLG-AB and BLG-AB bound with palmitic acid are 70.5° and 78.5° at pH 6.5, respectively. This result indicates fatty acids help to stabilise the BLG structure.

1.1.6.2 Retinol

The first report of retinol binding to BLG was in 1972 by Futterman and Heller (1972). Subsequently, the binding of retinol to BLG has been studied by a variety of methods, for example, fluorescence (Fugate and Song, 1980), equilibrium dialysis (Puyol *et al.*, 1991), affinity chromatography (Jang and Swaisgood, 1990), etc. The dissociation constant of retinol by different methods varies from 2.0×10^{-8} to 5.8×10^{-7}M.

From a solution study, Fugate and Song (1980) proposed that hydrophobic interactions play an important role in the binding of retinol to BLG and the nonpoplar part of retinol is mainly responsible for binding.

The binding affinity of retinol increases if chemically modifying BLG by methods such as methylation, ethylation (Dufour and Haertle, 1990a, b), esterification or alkylation (Dufour and Haertle, 1991). The increase of binding affinity may result from the increase in the overall hydrophobicity of BLG and therefore strengthen the binding interaction.

BLG from all species appears to bind retinol, but this ligand is not only found associated with BLG but also with other milk proteins such as serum albumin and α-lactalbumin (Puyol *et al.*, 1991) indicating that this interaction seems to be rather unspecific. Thus the functional relevance of retinol binding to BLG is also open to question.

The location of the retinol binding site on BLG and the functional significance of binding are still unclear. There are two divergent hypotheses on the retinol binding site.

(a) *The hypothesis of the internal calyx binding site:*

One hypothesis is from the modelling studies and proposes retinol is bound in the central calyx formed by the eight antiparallel β-strands (Papiz *et al.*, 1986; Sawyer *et al.*, 1998). The position of retinol was modelled directly from that in retinol-binding protein (RBP). In this model, the residues in the calyx make a variety of hydrophobic interactions with retinol, which is consistent with the result from solution binding studies which indicate that the binding of retinol to BLG mainly involves hydrophobic interaction (Fugate and Song, 1980; Jang and Swaisgood, 1990). Furthermore, the retinol hydroxyl group can make hydrogen bond contact with Lys69 and Trp19 can make close contact with the β-ionone ring of retinol.

The hypothesis of the internal calyx binding site was supported by the work of Chen *et al* (1993) and Cho *et al* (1994). Chen found that BLG with its α-helix removed by proteolysis still bound to retinol in a manner similar to the intact BLG. Cho (1994) found using site-directed mutagenesis that the residues of central calyx contribute to the binding interaction. In his experiment, four separate mutants, W19A, K70M, F136A and K141M were made where W19A or K70M mutated the residues in the central calyx while F136A or K141M mutated the residues in the external groove. It was shown that W19A and K70M mutants resulted in a dramatic decrease in the binding of retinoic acid compared to that of the F136A, K141M and wild-type BLG.

(b) *The hypothesis of the external groove binding site:*

The other hypothesis is based on a difference density map that shows some electron density found at a surface groove between the β-barrel and the α-helix and assigned to the retinol (Monaco *et al.*, 1987). In this model, the β-ionone ring of the retinol interacts with the side chain of Phe136 while the retinol hydroxyl group is hydrogen bonded to Lys141. However, since the electron density for retinol is very weak and ill-defined probably due to the poor data or low occupancy of the ligand and there is an error in the crystallographic analysis, the result is questionable.

Therefore, full identification of the binding site of retinol will require the complex structure to be solved by x-ray crystallography.

1.1.6.3 Other compounds

In addition to retinol and fatty acids, many other compounds also bind to BLG and most of them are hydrophobic in nature.

Many aliphatic and aromatic compounds, for example, butane, heptane, toluene and trifluorotoluene are reported to bind to BLG although the binding strengths are much weaker than those seen with retinol and fatty acids. These organic compounds are believed to bind to the same site that is located in the interior of the protein because this site is not affected by the state of dissociation from dimer to monomer (Papiz, 1982).

Some flavour compounds have been shown to bind to BLG and this property has been applied in the food industry for food additives. β-ionone (Dufour and Haertle, 1990b) and some alkanone flavours (O'Neill, 1987) are the examples.

β-ionone binds to BLG strongly with an association constant of $1.6*10^6(M^{-1})$.

Modifying BLG by chemical or enzymatic treatment results in increasing the binding affinity, which suggests changing the BLG conformation to alter its binding properties and consequently to broaden its possible applications in the food industry.

1.1.7 Conformational Change

As observed by ORD (Optical Rotatory Dispersion), BLG reveals three pH-dependant conformational transitions, which can be summarised as:

Conformational transition	Q	↔	N	↔	R	→	S
pH	4		6-6.5		7.8		> 8.5

The first transition assigned as Q↔N is reversible between pH=4 and 6 and is observed in bovine BLGA, B and C (Timasheff *et al.,* 1966; Mckenzie and Sawyer, 1967). This transition is characterised by an increase in sedimentation coefficient with an increase in pH. The increase in sedimentation coefficient seen is probably related to the contraction of the protein.

The second one, N↔R, is observed between pH 6 and 8 and is known as the Tanford transition (Tanford *et al.*, 1959). Because this conformational transition occurs in physiological conditions and is possibly associated with the biological function of

BLG, it has been the focus of many researches. This transition is characterised by a decrease in optical rotation ($[\alpha]_D$) from -25° at pH 6 to -48° at pH 8, indicating a change in the environment of an optically active group. It was observed by Tanford (Tanford *et al.*, 1959) that this transition results in a carboxyl group buried at pH below 7.0 becoming exposed at pH above 7.0. This transition is also detected by a decrease in the sedimentation coefficient from 3.2 to 2.6s at 20°C (Pederson, 1936), an increase in reactivity of a sulphydryl group with reagents containing Hg^{2+}(Dunnill and Green, 1965) and a change in the environment of tyrosines (Townend *et al.*, 1969). A decrease in the sedimentation coefficient is probably due to an expansion in the volume of BLG (Tanford *et al.*, 1959) or the dissociation of dimer to monomer (Mckenzie and Sawyer, 1967). Moreover, it was found recently that the thermal behaviour of BLG is also pH-dependent as a peak in the thermalgram (Qi *et al.*, 1995) disappears when the pH increases from 6.75 to 8.05.

After the project of this thesis started, the structures of BLG at pH 6.2, 7.1 and 8.2 have been solved by Qin (Qin *et al.*, 1998a) and reveal the structural basis of the behaviour of BLG during the Tanford transition. It has been shown that there is a distinct movement of a loop, the EF loop, as the pH is raised. The anomalous carboxyl group with pKa of 7.3 was probably identified as Glu89 by Brownlow *et al* (1997) but confirmed convincingly by Qin *et al* (1998a). At pH below 7.0, the buried Glu89 of the EF loop is hydrogen bonded to Ser116 of the GH loop whereas, at pH above 7.0, the hydrogen bond to Ser116 is broken and Glu89 is exposed along with the movement of the EF loop. The solvent accessible area of Glu89 is $4Å^2$ at pH 6.2, $116Å^2$ at pH 7.1 and $136Å^2$ at pH 8.2. A decrease in the sedimentation coefficient can be explained as an overall increase in molecular volume or surface area. The accessible surface area of BLG increases from $8240Å^2$ at pH 6.2 to $8460Å^2$ at pH 7.1 to $8670Å^2$ at pH 8.2 (Qin *et al.*, 1998a). An increase in reactivity of a sulphydryl group with a positively charged reagent is because the environment of Cys121 becomes more negatively charged and more susceptible to chemical modification. When the pH increases from 6.7 to 7.2, the residues (Lys101, Lys135, Arg148, Asp129, Asp137, Glu131 and Glu134) around Cys121 become more negatively charged and the average B factor of atoms with in 8Å of Cys121 increases from

16.2Å^2 to 30.0Å^2 (Qin *et al.*, 1998a). The increase in B-factor indicates these atoms are more unstable and consequently more susceptible to chemical modification. However, the structural study was unable to identify a dramatic change in the environment of tyrosines observed by Tanford as the conformations of the residues around these four tyrosines, Tyr20, 42, 99 and 102, are the same at pH 6.1, 7.1 and 8.2.

The last transition occurring above pH 8.5, assigned as R→S, is an alkalic denaturation. When the pH rises from 8 to 9, the dimer dissociates into monomer reversibly but the protein will denature and aggregate irreversibly above pH=9.0 (Townend *et al.*, 1960).

1.1.8 Sequence and structural homology within the proteins of lipocalin family

BLG is a member of the lipocalin family. The name of lipocalin, firstly suggested by Pervaiz and Brew (1987), comes from the Greek words "lipos" meaning fat and "calin" meaning cup. There are more than 20 proteins in this family as identified through sequence homology (Flower et al., 1996) with functions varying from insect camouflage to small hydrophobic molecule transport typified by serum retinol binding protein. The evolution of the lipocalin family from a common ancestor is supported by work on gene sequences, which share common intron/exon patterns among the members of this family (Ali and Clark, 1988). The crystal structures of several members have been reported, for example, retinol-binding protein (Newcomer *et al*, 1984; Cowan *et al*, 1990), BLG (Brownlow *et al.,* 1997), insecticyanin (Holden *et al*, 1987), bilin binding protein (Huber *et al*, 1987), odorant binding protein (Monaco *et al*, 1990), mouse major urinary protein, rat α 2u-globulin (Bocskei *et al.*, 1992) and retinoic acid binding protein (Newcomer, 1993). The crystal structures so far available reveal the typical lipocalin to be an 8-stranded antiparallel β-barrels arranged to form a conical central calyx or cavity in which the hydrophobic ligand is located (Newcomer *et al.*, 1984). There is an α-helix on the outer surface of the β-barrel and the amino acid sequence contains 3 structurally conserved regions (SCR's or sequence motifs) together with one or more disulphide

bridges. The calycin super-family extends the family to include proteins like the fatty acid binding proteins (FABP) which do not contain the second of these 3 sequence patterns but which are also antiparallel β-sheet proteins, though with 10 β-strands rather than 8. However, in common with their distant lipocalin relatives, their binding site is also within a central cavity.

Although there is low overall sequence identity throughout the lipocalin family, typically as low as 10 or 20%, there are some regions where the sequence conservation is very much greater. The three structurally conserved regions (SCRs) found in lipocalins (North, 1989) are all located close to each other at the bottom (closed end) of the β-barrel. It has been proposed that these structurally conserved regions probably consist of a receptor binding site (Yewdall, 1988; North, 1989). These three regions are discussed below:

(a) The first SCR is located on strand A with a conserved sequence of h-a-x-x-b-h-x-Gly-x-Trp-y-x-h-h around residue 20.

Where a is usually an acidic residue, b is usually a basic residue, h is a hydrophobic residue, y is aromatic residue and x is any kind of residue.

(b) The second SCR is located on the FG loop with a conserved sequence of h-h-x-Thr-Asp-Tyr-x-x-y-h around residue 100. It is very common feature that the main-chain in this region adopts an unfavourable conformation and is stabilised by specific sidechain-mainchain hydrogen bond from conserved threonines and aspartates. For example in BLG, Tyr99 is in the disallowed region of the Ramachandran plot in all BLG structure determined to date. There are also four mainchain-sidechain hydrogen bonds formed in this region involved Asp98, Tyr99 and Lys100 (Brownlow *et al.*, 1997). Moreover, Tyr99 adopts a classic γ-turn conformation that is common to nearly all of the lipocalin structures. It has been suggested this γ-turn may be involved in some biological functions (North, 1989)

(c) The third SCR is located on the loop before α-helix with a conserved basic residue at Arg124. Arg124 is at the centre of these three SCR regions and the side chain of Arg124 lies parallel to the side chain of Trp19, which is in the first SCR region.

1.1.9 The thermal denaturation of BLG

The effects on bovine milk of high temperature are very important in diary science research as the processing of milk, for example, pasteurization, sterilisation and dehydration, usually involves the use of high temperature processes.

The heat stability of BLG has been studied for a long time as BLG is involved in the formation of fouling deposit from heated milk. Moreover, the binding of BLG to casein upon denaturation increases the yield of cheese by increasing the amount of protein. These and various other reactions must be well characterised in order to improve the production of diary products without loss of an acceptable flavour.

Several methods have been used to study the thermal stability of BLG, for example, optical rotatory dispersion (Sawyer *et al.*, 1971), differential scanning calorimetry (DSC) (Park and Lund, 1984), CD (Lapanje and Poklar, 1989). The most popular method is differential scanning calorimetry (DSC).

BLG-A or BLG-B contains one cysteine and two disulphide bridges per monomer. It has been suggested these cysteines are able to undergo sulphydryl oxidation of disulphide interchange interaction (-SH/SS). These reactions result in the denaturation of BLG. Briggs and Hull (1945) suggested there are two distinct stages involved in the denaturation of BLG. The first step is dissociating dimer into monomer and unfolding the polypeptide chain to a random-coiled conformation. The second step is the unfolded protein undergoes -SH/SS exchange reactions to yield irreversible aggregates.

The thermal behaviour of BLG in the presence of small molecule ligands and other milk components has been investigated. Puyol *et al* (1994) have found binding of palmitic acid to BLG increased significantly the denaturation temperature as mentioned in section 1.1.6.1. The increase in the resistance to heat denaturation indicates the binding of palmitic acid to BLG stabilised BLG structure. This phenomenon is similar to the stabilising effect of other ligand bound to other milk protein, e.g. fatty acid bound to bovine serum albumin (BSA) (Gumpen *et al.*, 1979) or Ca bound to α-lactalbumin (Relkin *et al.*, 1993).

It is also found lactose stabilises BLG against thermal denaturation (Park and Lund, 1984).

However, for BLG with bound retinol, the denaturation temperature (74.5°) is lower than that of BLG with bound palmitic acid although the denaturation temperature is still higher than that of native BLG (70.5°). This result indicates the binding the retinol to BLG is probably a less important factor in the stabilising of BLG structure compared to palmitic acid.

1.2 CYP3

1.2.1 Introduction

Cyclophilins are omnipresent, highly conserved proteins found in diverse prokaryotic and eukaryotic organisms (Galat and Metcalfe, 1995). This protein has been found in a variety of species ranging from bacteria to vertebrates, including human, plant, and several parasitic species. The biological functions of cyclophilins include high affinity binding to the immunosuppressant cyclosporin (CsA), possession of the peptidyl-prolyl *cis-trans* isomerase activity (PPIase activity) and assisting in the folding of several proteins in the cell.

Cyclophilin was first identified and purified from bovine thymus and human spleen by Handschumacher *et al* (1984) and Harding *et al* (1986), respectively. This protein has a high affinity for CsA with a K_d of 46nM by tryptophan fluorescence (Liu et al., 1990). CsA is a cyclic undecapeptide and is the principal immunosuppressant drug used to prevent organ rejection after transplant operations (Beveridge and Calne, 1995). CsA also has therapeutic value in the treatment of autoimmune disorders, including uveitis (Nussenblatt *et al.*, 1983), type I diabetes (Stiller *et al.*, 1984) and encephalomyelitis (Hinrichs *et al.*, 1983). Additionally, it has been reported that sub-immunosuppressive levels of CsA have potent anti-parasitic activity on a wide range of both helminth and protozoan parasites (Page *et al.*, 1995a,b; Bell *et al.*, 1996). Moreover, non-immunosuppressive CsA analogues have been found to have an effect on the malaria parasite (Bell *et al.*, 1994). Recently, CsA has been shown to inhibit the association of cyclophilin A and HIV-1 virions, consequently reducing the HIV-1 virion infectivity, as the binding of cyclophilin A to HIV plays an important role in HIV-1 infectivity (Thali *et al.*, 1994).

Cyclophilins also have peptidyl-prolyl *cis-trans* isomerase activity (PPIase activity). In independent studies, an enzyme isolated by Fischer *et al* (1984) and Lang *et al* (1987) catalysed the *cis→tran* isomerization of proline-containing peptide bonds. This enzymatic activity was therefore named as peptidyl-prolyl *cis-trans* isomerase (PPIase) activity. The first PPIase was purified and sequenced by Fischer *et al* (1989a) and revealed that the PPIase isolated from pig kidney shared identity with the cyclophilin from bovine thymus.

Therefore, cyclophilins are not only the receptors for the immunosuppressant drug, CsA, but are also PPIases. Since the PPIase activity of cyclophilins can be completely inhibited by CsA, it has been proposed that the binding site of CsA is identical to the PPIase active site. This proposal has been confirmed by crystallographic studies as both CsA and proline-containing peptides bind in the same place, at the surface of the protein (Kallen *et al.*, 1991; Pflugl *et al.*, 1993).

Since the first structure, a tetrapeptide substrate, acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin (acAAPAamc) bound with human cyclophilin (hCYPA), was solved by Kallen in 1991 (Kallen *et al.*, 1991), a number of structures of cyclophilin complexed with different ligands have been solved in the past few years. All ligands bind in a long and deep groove located on the surface of the protein. The liganded structures will be discussed in sections 1.2.4 and 1.2.5. These liganded structures could provide a starting point to reveal the immunosuppressive reaction and understand the *cis→tran* isomerization.

Cyclophilin usually exists as many distinct isoforms in a single species. For example, there are at least four different cyclophilin isoforms in human (Galat and Metcalfe, 1995). Human cyclophilin A (hCYPA) consisting of 165 amino acids (17800 Da) is the most abundant and well-characterised member in the cyclophilin family (Handschumacher *et al.*, 1984; Harding *et al.*, 1986). Human cyclophilin B (hCYPB) has 208 residues (22742 Da) and shares a sequence homology of 64% with hCYPA (Price *et al.*, 1991; Hasel *et al.*, 1991). Human cyclophilin C (hCYPC) has 212 residues with molecular weight of 22795 Da (Thalhammer *et al.*, 1992) while human cyclophilin D (hCYPD) has 207 residues (Bergsma *et al.*, 1991). These isoforms show great similarity in overall folding and have a highly conserved CsA-binding

domain (Taylor *et al.*, 1997). Significant divergence occurs in the N- and C-terminal regions. Compared to hCYPA, hCYPB and hCYPC contain an extra signal domain at the N-terminus that directs them to the endoplasmic reticulum (ER). Human cyclophilin D also has a signal domain at the N-terminus that directs it to mitochondria. Moreover, a large cyclophilin, human cyclophilin 40 (hCYP40), with a molecular weight about 40 kDa has been isolated and cloned from a human cell line (Hoffmann *et al.*, 1995a,b). Human CYP40 contains two domains; the N-terminal cyclophilin-like domain reveals 40% sequence identify with hCYPA.

Recently, 11 cyclophilin isoforms (Page *et al.*, 1996) have been cloned and sequenced from the free-living nematode *Caenorhabditis elegans* (C. *elegans).* One of these C. *elegans* cyclophilins, cyclophilin 3 (CYP3) is the main protein to be discussed in this thesis. It contains 173 residues with molecular weight of 18550Ds. Based on the sequence alignment, CYP3 reveals 67% sequence homology with hCYPA. The overall structure of CYP3 is very similar to that of hCYPA with the exception of the extra loop containing 7 inserted residues in CYP3. The structure and biological characteristics of CYP3 will be discussed in chapter 4.

1.2.2 The family of *Caenorhabditis elegans* cyclophilins

All of the eleven isoforms of cyclophilins from *C.elegans* have been cloned and expressed in *Escherichia coli*, except cyclophilin 7 (CYP7). CYP7 was excluded due to its extremely high identity to cyclophilin 3 (CYP3); 87% identify over the entire sequence and 100% identity over the CsA-binding domain. The sequence alignment reveals that 7 of 11 of these isoforms in *C.elegans* contain a 7-8 amino acid insert compared to hCYPA. The important features and possible functions of this insert are discussed in section 5.1.3.

CYP1, 3, 5, 6, 7 have a conserved CsA binding domain (CBD) compared to CYPA while CYP2, 4, 8, 9, 10 have one to four mutant residues in the CBD. The CsA binding domain refers to the residues binding to CsA in human CYPA as will be described in section 1.2.5. Based on the sequence comparison, CYP2, 3, and 7 most closely resemble CYPA and are thus classified as CYPA-like isoforms.

These isoforms reveal great divergence in the N-terminal and C-terminal regions. For example, CYP1, 5 and 6 possess a signal domain in the N-terminal region. The 25 amino acid hydrophobic signal peptide found in CYP6 is similar to that seen in human cyclophilin B and this protein is classified as a CYPB-like isoform. CYP1 and CYP5 also contain a signal peptide at the N-terminus. However, these signal peptides are more similar to human cyclophilin C, and hence classified as CYPC-like isoforms. CYP4 is actually a three domain protein. These three domains consist of a N-terminal domain (254 amino acids), a cyclophilin domain and a C-terminal domain (220 amino acids). CYP8 and CYP9 contain two domains. One is a cyclophilin domain and the other is an extended C-terminal domain. CYP4, 8 and 9 together with CYP10 and CYP11 are classified as divergent cyclophilin isoform, or cyclophilin D like isoforms. The similarity and difference between these 11 isoforms are summarised in Table 1-D.

Table 1-D: The family of C.elegans cyclophilins

	Size (aa)	CsA-binding domain	Signal peptide	Inserted residues	Variable N-terminal domain	Variable C-terminal domain	pI	Isoform -like
CYP1	192	conserved	yes (19aa)	yes	no	no	6.7	C
CYP2	171	A103C	no	yes	no	no	8.1	A
CYP3	173	conserved	no	yes	no	no	8.8	A
CYP4	524	T73H, A103K, W121Y	no	no	yes	yes	7.9	D
CYP5	204	conserved	yes (28aa)	no	no	yes	9.5	C
CYP6	201	conserved	yes (25aa)	no	no	yes	5.8	B
CYP7	171	conserved	no	yes	no	no	8.5	A
CYP8	466	A103R, W121H	no	yes	no	yes	10.7	D
CYP9	309	T73R, A103K, W121H, L122C	no	yes	no	yes	8.6	D
CYP10	147	T73K, A103N, W121H, H126Y	no	no	no	no	5.3	D
CYP11	183	W121F	no	yes	no	no	6.5	D

Key:

- aa: amino acid
- size: the sizes of proteins include signal peptide, the extended N- and C-terminal domain where applicable
- CsA-binding domain: the residues in the human CYPA which bind CsA, with changes in this conserve domain represented in single-letter amino acid code and numbering referring to the position in the human CYPA
- inserted residues: a central 7-8 amino acid insert
- isoform-like: A (cyclophilin A-like isoform), B (cyclophilin B-like isoform), C (cyclophilin C-like isoform) D (cyclophilin D-like isoform)

CYP3 is the subject of part of this thesis (see chapter 5). It contains 173 amino acids (18550Da) with a pI of 8.8. It is one of the most abundantly expressed isoforms. Based on sequence alignment, this protein has the conserved CsA binding domain. There are no signal peptides at the N-terminus and there are no extended N-terminal and C-terminal domains. However, compared to hCYPA, CYP3 contained 7 inserted residues starting from Lys48 in the CYP3 sequence.

1.2.3 Overall structure of human cyclophilin A (hCYPA)

The structure of cyclophilin A contains 8 antiparalleled β-strands and 2 α-helices (Kallen *et al.*, 1991; Ke *et al.*, 1991). The overall shape of the structure is globular with a dimension of approximately 34x33x30 ($Å^3$). These 8 antiparalled β-strands form a β-barrel and 2 α-helices sit on the top and bottom of the β-barrel (Figure 1-2). Cyclophilins share a common feature, "a β-barrel formed by antiparallel β-strands", with the lipocalin family. However, cyclophilin is unique and distinct from the lipocalin family in terms of topology and overall morphology of the structure.
The β-barrel of cyclophilin has a +1, -3, -1, -2, +1, -2, +3 topology, which is different from the $(+1)_n$ topology found in the lipocalin family or the (-3, +1, +1) Greek key topology frequently found in β-strand-predominated structures (Ke *et al.*, 1992).
Inside the barrel of cyclophilin, a hydrophobic core contains most of the hydrophobic residues. Other hydrophobic residues are found in the CsA binding site and the contact region of the β-barrel and 2 α-helices. The residues contributing to the hydrophobic core include seven aromatic residues (Phe8, Phe22, Phe36, Tyr48, Phe53, Phe112 and Phe129) and eight hydrophobic residues (Val6, Val20, Leu24, Leu39, Ile56, Leu98, Met100 and Ile114).
The major difference between cyclophilins and the lipocalin proteins is that no ligand or substrate can bind to the hydrophobic core of the cyclophilins while the ligands do bind inside the hydrophobic core of the lipocalins as discussed in section 1.1.8.

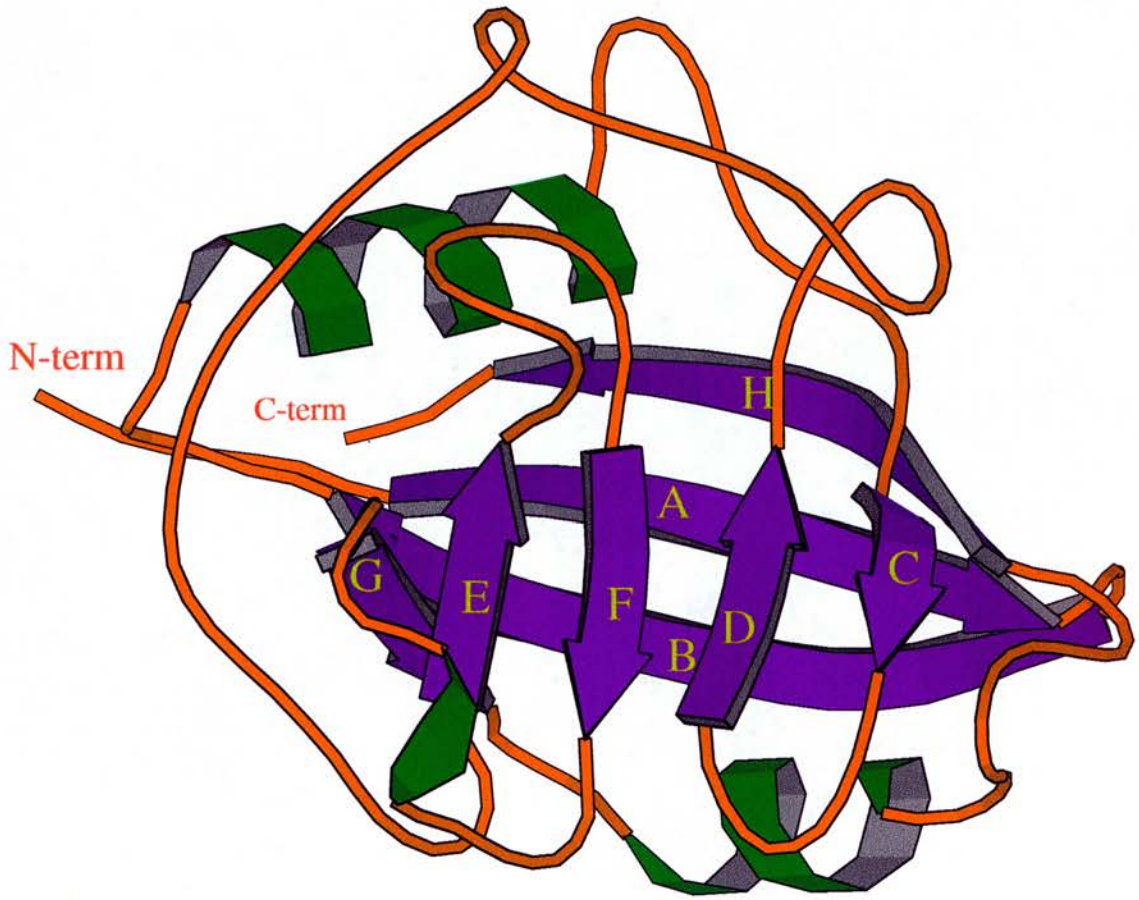


Figure 1-2.Diagram of the structure of hCYPA, produced by MOLSCRIPT (Kraulis, 1991). The eight labelled beta-strands are shown in purple, the two helices are shown in green and the joining loops are shown in orange.

The reason why the hydrophobic core of cyclophilin is not accessible to the ligands is because the barrel of cyclophilin is closed by the two helices, which block the entrance to the core and consequently hinder the ligands to enter the core.
Moreover, inside the core, the hydrophobic residues pack tightly and consequently exclude the possibility of ligand binding within the barrel. Actually, based on the structural studies which will be discussed in section 1.2.4 and 1.2.5 of cyclophilin complexed with a variety of ligands, all ligands bind in a long and deep groove located on the surface of the protein between one face of the β-barrel and the FG loop.
As can be seen, the binding sites of two proteins are significantly different due to different arrangement and surroundings of the binding sites although they share the similar motif.

1.2.4 Cyclophilin-ligand complex structures

Since the first complexed structure, a tetrapeptide substrate (acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin) bound to hCYPA was solved by Kallen in 1991 (Kallen *et al.*, 1991), many studies have concentrated on the crystallographic study of cyclophilin in order to understand the immunosuppressive reaction and the *cis-trans* isomerization mechanism. As most complexed structures were studied with hCYPA and the major protein studied in this thesis, CYP3, is a hCYPA-like protein in terms of sequence homology and structure, the reviews of the complexed structures discussed in sections 1.2.4 and 1.2.5 are therefore focused on the hCYPA although several complexed structures with hCYPB (Mikol *et al.,* 1994), murine CYPC (Ke *et al.,* 1993b) and *E.coli* CYPA (Konno *et al.,* 1996) have been published. Many complexed structures that have been reported, include various proline-containing peptides (Ke *et al.,* 1993a; Zhao *et al.*, 1996a,b), CsA (Mikol *et al.*, 1993; Pflugl *et al.*, 1993) and CsA derivatives (Kallen *et al.*, 1998). Recently, the structures of hCYPA complexed with a 25 amino acid fragment of HIV-1 gag protein (Zhao *et al.,* 1997) and hCYPA complexed with N-terminal domain (151 residues) of the HIV-1 capsid protein have been solved (Gamble *et al.*, 1996).
The binding site is located on the surface of the protein and formed by the strands C, D and F (see Figure1-3). The detail of these complexed structures will be discussed later in this section and in section 1.2.5. The close contact and hydrogen bond distances between cyclophilin and a variety of substrates are summarised in Table 1-E.

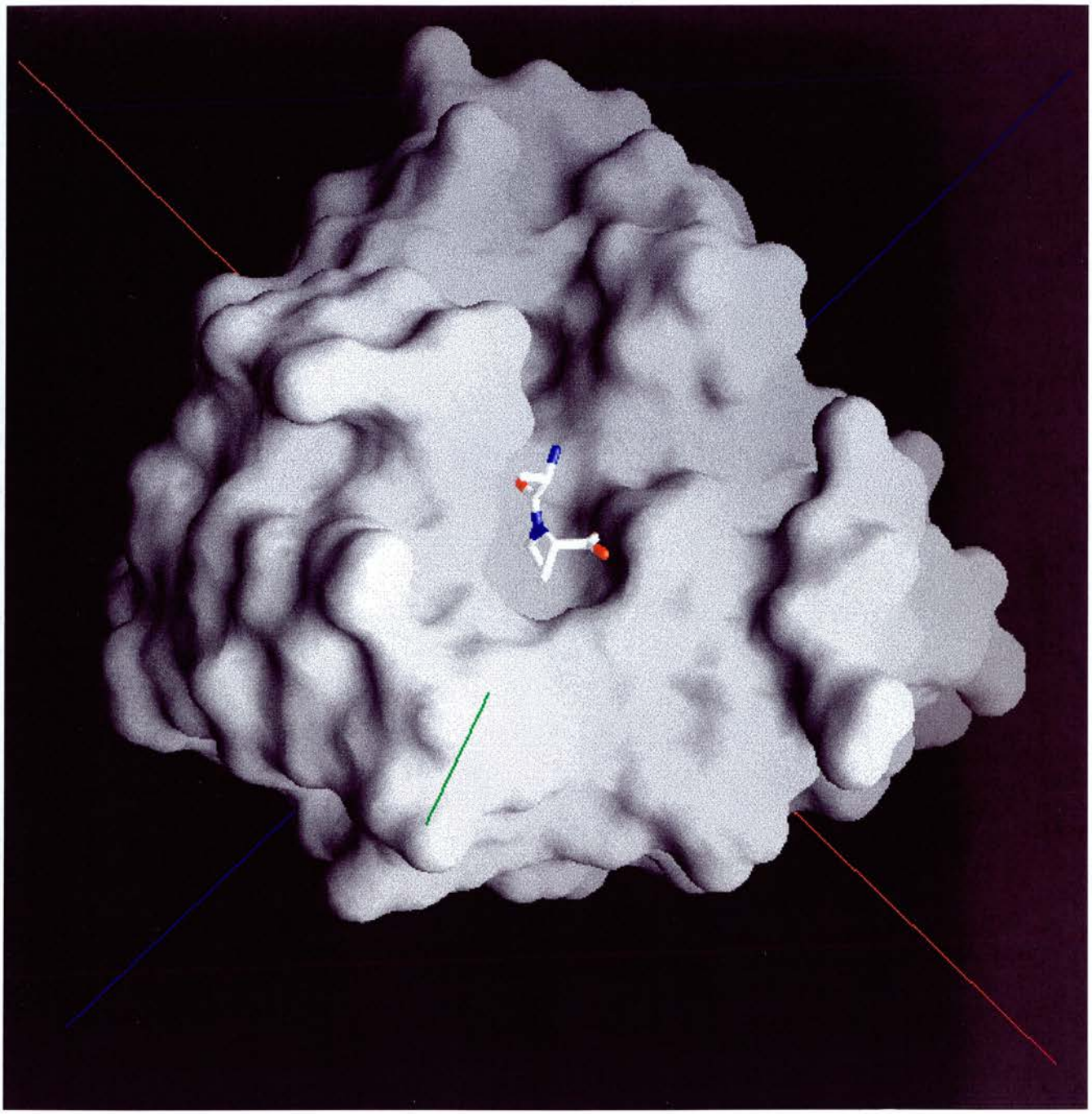


Figure 1–3.Diagram of the molecular surface of hCYPA, prepared by GRASP (Nicholls et al., 1991).The orientation of the model is similar to that of Figure 1–2. The binding site is presented at the centre of the diagram with the model of Ala–Pro to highlight the hydrophobic pocket.

Table 1-E: Cyclophilin-ligand close contacts and hydrogen bond distance (Å)

	acAAPAamc	sAAPFna	X-P (X=S, H, A, G)	CsA	25mer	HIV-1 CA
O/Asn71					ND1/His87 (2.9)	ND1/His87 (3.6)
O/Gly72					N/Ala88 (2.9)	N/Ala88 (2.9)
NE2/Gln63		o/Suc (2.8)	N/Xaa (2.7-2.9)	O/MeBmt (3.2)	O/Ala88 (3.2)	O/Ala88 (3.2)
OE1/Gln63			O/Pro3 (3.0-3.1)			
O/Asn102	N/Ala2 (3.25)	N/Ala2(3.1)	N/Xaa (2.7-2.9)	N/Abu2 (2.9)	N/Gly89 (2.8)	N/Gly89 (2.9)
N/Asn102	O/Ala (3.35)		O/Xaa (3.0-3.1)	O/MeVal11 (3.5)		
NH1/Arg55	O/Pro3 (2.9)	O/Pro3 (2.65)	OT/Pro2 (2.7-2.8)	O/MeLeu10 (2.8)	O/Pro90 (2.7)	O/Pro90 (2.9)
NH2/Arg55		O/Pro3 (3.15)		O/MeLeu10 (2.9)	O/Pro90 (2.7) OE1/Gln95 (2.9)	O/Pro90 (2.7) OE1/Gln95 (3.7)
NE/Arg55			O/Pro2 (2.8-2.9)			
NE1/Trp121		O/Phe4 (3.0)		O/MeLeu9 (2.9)	O/Ile91 (3.1)	O/Ile91 (3.3)
NH1/Arg148		OD1/na (2.75)				
CE1/His126	O/Ala2 (3.1)					
CE1/Phe60	amc					

Key:
(1) Numbers in parentheses refer to the hydrogen bond distance
(2) Distances from the cyclophilin-dipeptide structures Ser-Pro, His-Pro, Ala-Pro and Gly-Pro are given as ranges

(a) *The structure of hCYPA complexed with dipeptides*:

Several complexed structures of dipeptide bound to CYPA have been determined and refined at between 1.64Å and 2.1Å. The first structure reported was the Ala-Pro binding to hCYPA by Ke (Ke *et al.*, 1993a). The other three complexed structures, Ser-Pro/hCYPA, His-Pro/hCYPA and Gly-Pro/hCYPA, were solved in 1996 by Zhao (Zhao *et al.*, 1996b). Comparison of these four structures shows proline-containing dipeptides bind to cyclophilin in the *cis* form with an omega angle of between 5° and -4°. The structure of Ala-Pro binding to hCYPA is shown in Figure 1-4 to illustrate the interactions between dipeptide and cyclophilin.

Proline makes hydrophobic contact with the hydrophobic pocket of CYPA, which is composed of Phe60, Met61, Phe113, Ile122 and His126. The two C-terminal oxygens form three hydrogen bonds with Arg55 and Glu63. The N-terminal nitrogen forms a hydrogen bond with the carbonyl oxygen of Asn102 while the carbonyl oxygen forms a hydrogen bond with the NH nitrogen of Asn102. The side chains of the Xaa residue in the dipeptide of Xaa-Pro do not make strong contacts with cyclophilin. The weak interaction of the Xaa residue with CYPA has indicated their minor effect on the *cis-trans* isomerization and therefore explains the lack of catalytic specificity of cyclophilin.

(b) *The structure of hCYPA complexed with tetrapeptides*:

Two structures of human recombinant cyclophilin A complexed with a tetrapeptide substrate have been reported. The structure of hCYPA complexed with acetyl-Ala-Ala-Pro-Ala-amidomethylcoumarin was the first cyclophilin structure to be solved (Kallen *et al.*, 1991). The other is the structure of hCYPA complexed with succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (sAAPFna) (Zhao *et al.*, 1996a). sAAPFna is a substrate commonly used for the assay of peptidyl-prolyl *cis-trans* isomerization of both cyclophilin and FK506 binding protein (FKBP). Both structures show the tetrapeptides are bound to CYPA in the *cis* form. The structure of sAAPFna binding to hCYPA is shown in Figure1-5 to illustrate the interactions between tetrapeptide and cyclophilin.

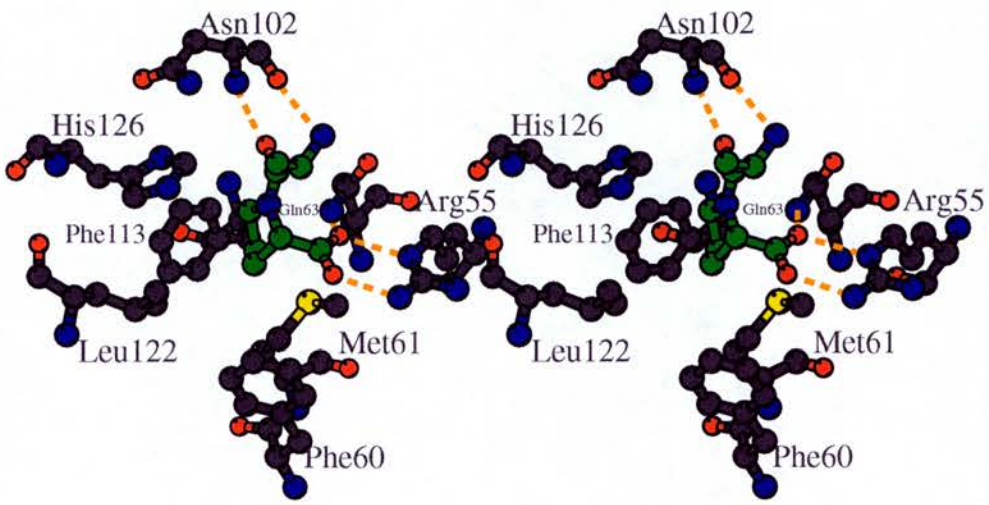


Figure 1-4.Stereodiagram of the binding site of hCYPA with bound Ala-Pro, produced by MOLSCRIPT (Kraulis, 1991). The residues forming the hydrophobic pocket are labelled. Some of the hydrogen bonds are marked by broken lines. The carbon atoms of Ala-Pro are shown in green

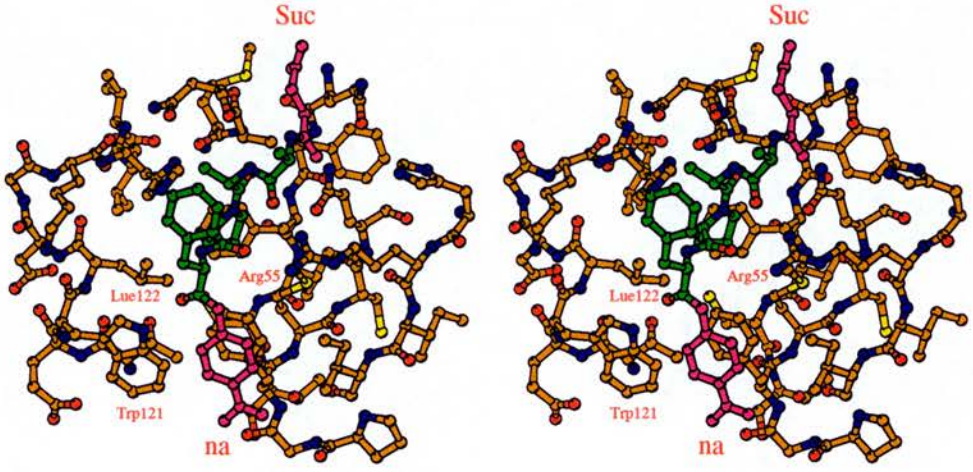


Figure 1-5.Stereodiagram of the binding site of hCYPA with bound sAAPFna, produced by MOLSCRIPT (Kraulis, 1991). The succinyl (Suc) and nitroanilide (na) groups of sAAPFna are labelled and shown in magenta. The carbon atoms of sAAPFna are shown in green.

The *cis* proline makes hydrophobic contact with the hydrophobic pocket of CYPA. The N-terminal nitrogen forms a hydrogen bond with the carbonyl oxygen of Asn102 while the carbonyl oxygen forms a hydrogen bond with the NH nitrogen of Asn102. The other hydrogen bond is formed by the prolyl carbonyl oxygen with the side chain of Arg55. These three hydrogen bonds are conserved in most structures of CYPA complexed with a variety of substrates (refer to Table1-E). Comparison of the structures of tetrapeptides bound to cyclophilin and that of dipeptides bound to cyclophilin reveals the major difference is the binding of Arg55. Arg55 adopts different conformations in these two types of complexed structures. In the tetrapeptide complexed structures, Arg55 adopts an identical conformation to that of the native structure. However, the torsion angle around the CD-NE bond of the side chain of Arg55 rotates from -140° in the tetrapeptide complexed structure to 70° in the dipeptide complexed structure. The other difference between these two complexed structures is the absence of a water molecule at the binding site in the tetrapeptide complexed structure. The water is hydrogen bonded to the side chain of Gln63 in the dipeptide complexed structure and was proposed to assist the *cis-trans* isomerization in the "solvent-assisted mechanism". The details of "solvent-assistant mechanism" will be discussed in section 1.2.6.

(c) *The structure of hCYPA complexed with HIV-1 capsid protein*:

The structure of hCYPA complexed with the N-terminal domain (residues 1 to 151) of the HIV-1 capsid protein (CA_{151}) was solved to 2.30Å by Gamble in 1996 (Gamble *et al.*, 1996). At the same time, the structure of a 25 amino acid peptide (residues 81 to 105) of the HIV-1 capsid protein (25-mer) bound with recombinant hCYPA was solved independently to a resolution of 1.8Å by Zhao (Zhao *et al.*, 1997) with 12 of the 25 amino acids (Val86-Arg97) well-fitted into the density map while the rest of the residues are disordered. Both structures show the capsid protein binds at the active site on the surface of the cyclophilin. There are no large conformation changes seen in either the cyclophilin or the capsid protein once the complex forms. CA_{151} is a flat molecule with overall dimensions of 18Å x 35Å x 40Å and contains 7 helices. The loop that connects helix 4 and 5 binds to the cyclophilin in an extended conformation (see Figure 1-6).

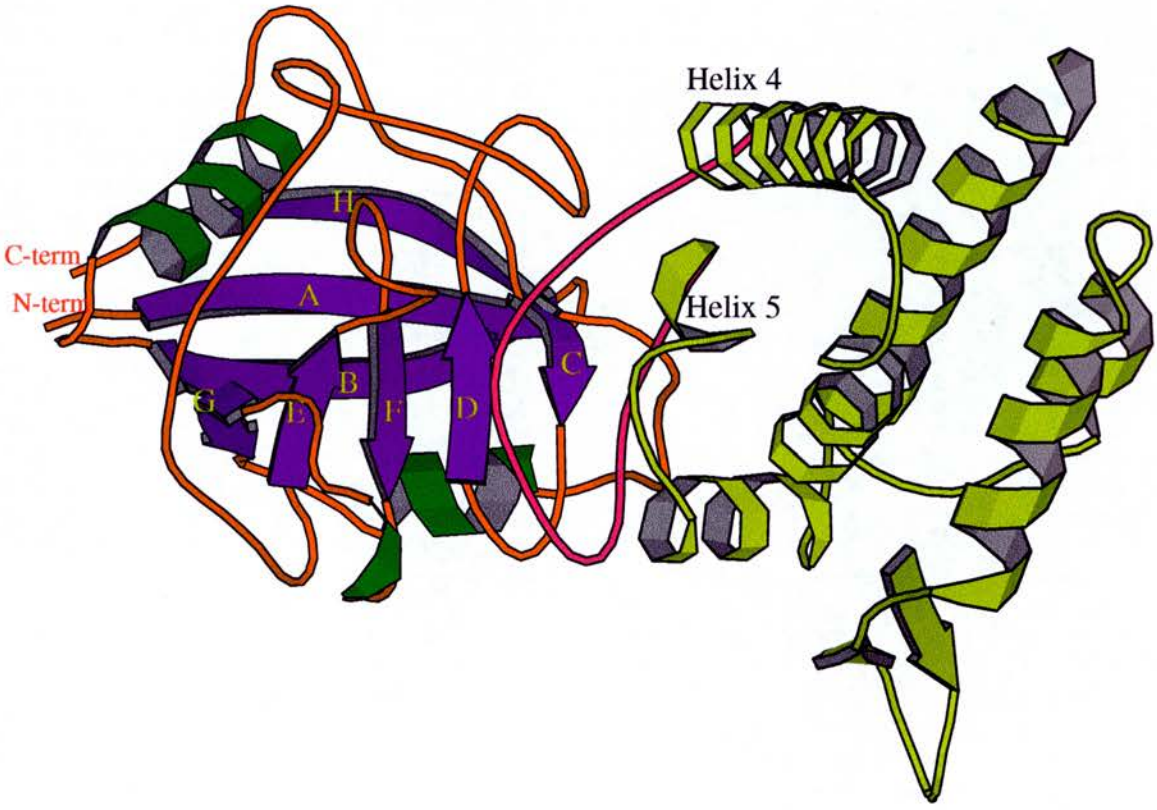


Figure 1-6.Ribbon diagram of the structure of hCYPA complexed with CA151 produced by MOLSCRIPT (Kraulis, 1991). The eight labelled strands of hCYA are shown in purple, the two helices are shown in green and the joining loops are shown in orange. The structure of CA151 are shown in lime green. The loop (magenta) that connects helix 4 and 5 binds to cyclophilin in the binding site.

Residues 85 to 93 of the capsid protein are involved in numerous hydrophobic interactions and eight hydrogen bond interactions with hCYPA (refer to Table 1-E). Pro90 of the capsid protein makes strong hydrophobic contacts with the residues in the hydrophobic pocket of CYPA. However, Pro90 adopts the *trans* conformation, in contrast to the *cis* conformation observed in other CYPA-peptide structures. In addition, Gly89, the residue preceding Pro90, has a unfavourable backbone conformation with ϕ=133° and ψ=160°. This unusual backbone conformation may be adopted by Gly whereas it is inaccessible to other amino acids because of the stereochemical constraints imposed by their side chains. Site-directed mutagenesis has shown Gly89 is essential for binding as mutation of Gly89 to alanine inhibits the CYPA-gag interaction (Colgan *et al.*, 1996; Braaten *et al.*, 1996). It has been proposed that cyclophilin plays a role as a chaperone or an essential functional component instead of as a *cis-trans* isomerase in its interactions with HIV-1. The reasons are as follows:

Firstly, the capsid protein binds to hCYPA in the *trans* form while the other proline-containing peptides bind to CYPA in the *cis* form. Secondly, only Pro90 binds to CYPA although there are three Proline residues (Pro85, Pro90 and Pro93) in the binding interface between capsid protein and hCYPA. If cyclophilin is essential in the folding of HIV-1 gag protein, it might be expected these three prolines would bind equally to the active sites of CYPA. However, the complexed structure shows only Pro90 binds to CYPA. Finally, a glycine residue preceding proline is essential for the interaction between cyclophilin and gag protein while there is no specificity in the requirement of the residue preceding proline in the other CYPA-peptide complexes. This result indicates the binding of capsid protein and CYPA is sequence specific.

In conclusion, the *cis-trans* isomerization activity of hCYPA may not be involved in the interactions between cyclophilin and HIV-1 although whether Gly89-Pro90 can undergo *cis-trans* isomerization by cyclophilin is still under investigation.

1.2.5 The structures of CsA and CsA derivatives bound to hCYPA

(a)*The structure of hCYPA complexed with CsA*:

Five structures of hCYPA complexed with CsA have been reported. Three were solved by X-ray crystallography, each with a different crystal form (Mikol *et al.*, 1993; Pflugl *et al.*, 1993; Ke *et al.*, 1994) while the other two were determined by NMR in solution (Wuthrich *et al.*, 1991; Weber *et al.*, 1991; Spitzfaden *et al.*, 1992; Theriault *et al.*, 1993). All these structures show CsA binds at the PPIase active site, which is located at the surface of cyclophilin (see Figure 1-7).

13 residues (Arg55, Phe60, Met61, Gln63, Gly72, Ala101, Asn102, Ala103, Glu111, Phe113, Trp121, Leu122 and His126) form the CsA binding pocket and are within 4Å of the bound CsA (see Figure 1-8).

Arg55, Ans102 and Gln63 make five hydrogen bond contacts with CsA. In addition, five rigid water molecules mediate hydrogen bond interactions between CsA and CYP. CsA is a cyclic undecapeptide which has 7 of the 11 peptides N-methylated (see Figure 1-9).

Analysis of the hCYPA-CsA complex structure shows residues 1, 2, 3, 9, 10 and 11 of the CsA are in contact with hCYPA (see Figure 1-8). MeVal11 makes the most strong contact with hCYPA although residues 1, 2, 3, 9 and 10 also contribute to binding. When compared to the structures of proline-containing peptides bound to CYPA, MeVal11 occupies the position usually taken by proline in the binding pocket. The remaining residues, 4, 5, 6, 7 and 8 are exposed on the CYPA surface (see Figure 1-8) and are proposed to have specific interactions with calcineurin (Liu *et al.*, 1992), a key protein in the T cell activation pathway. The role of MeVal11 in binding to cyclophilin is supported by site-directed mutagenesis studies. Mutation of MeVal11 to MeAla11 or MeLeu11 resulted in a decrease in cyclophilin binding and immunosuppressive activity (Mikol *et al.*, 1993). CsA adopts a significantly different conformation when bound to hCYPA. Free CsA in solution has a tightly folded conformation with four intramolecular hydrogen bonds (Loosli *et al.*, 1985).

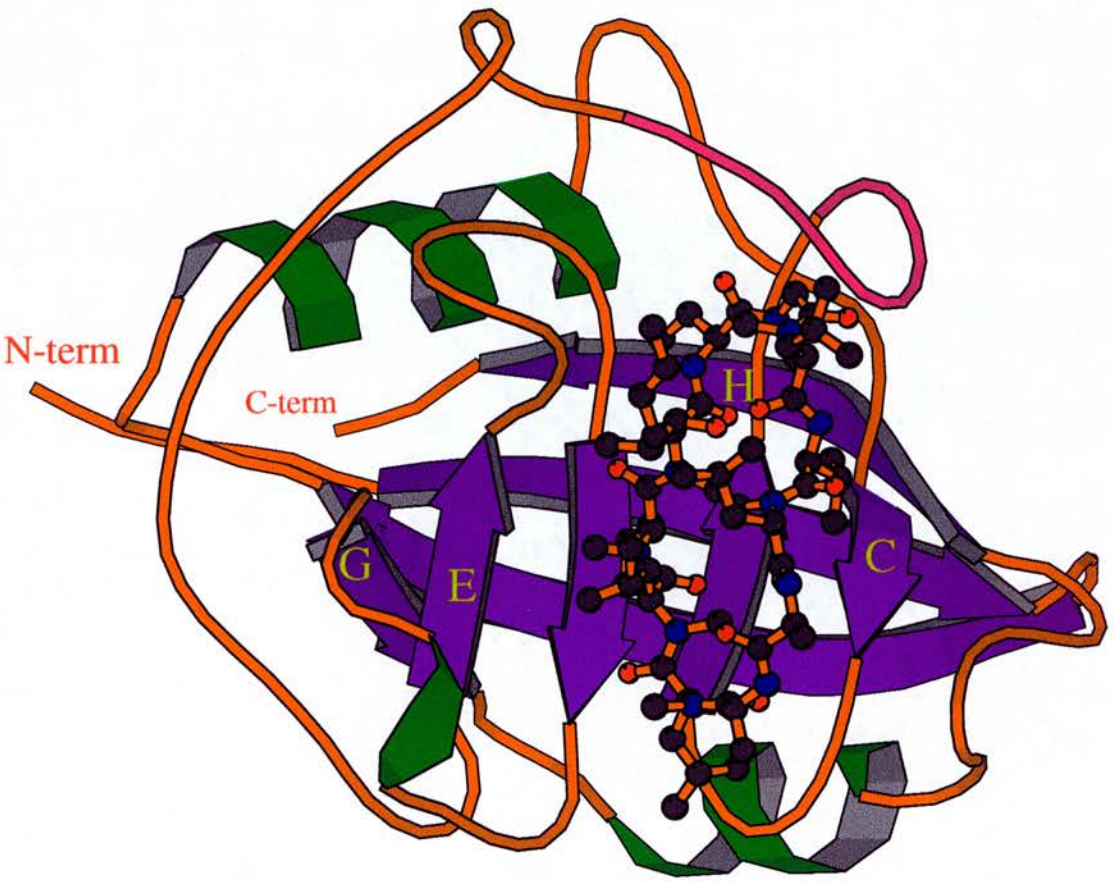


Figure 1-7.Diagram of the structure of hCYPA with bound CsA,produced by MOLSCRIPT (Kraulis, 1991). The eight labelled strands of hCYPA are shown in purple, the two helices are shown in green and the joining loops are shown in orange. The 70s loop (residues 67 to 76) is shown in magenta. The carbon atoms of CsA are shown in black.

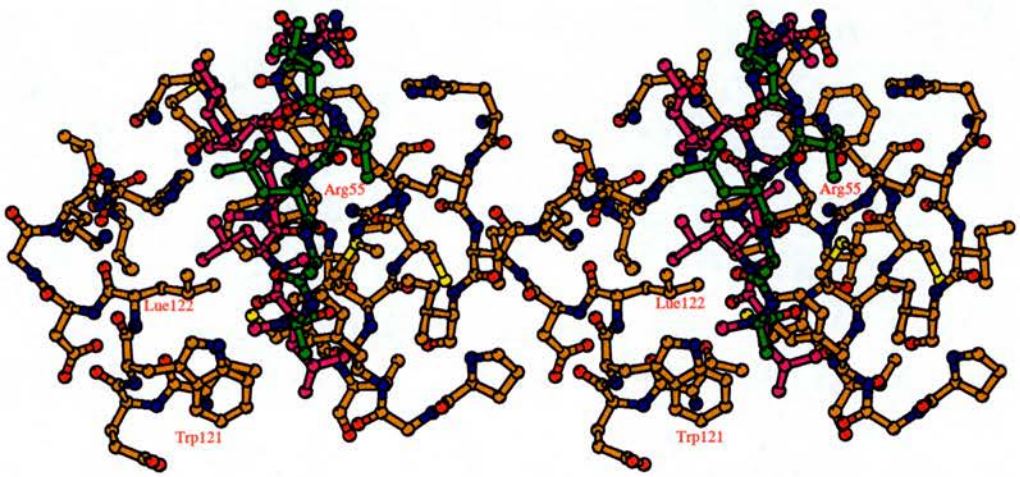


Figure 1-8.Stereodiagram of the CsA docked into the binding site of hCYPA. The carbon atoms of residues 1, 2, 3, 9, 10 and 11 of CsA, which are in contact with hCYPA, are shown in magenta while the carbon atoms of residues 4, 5, 6, 7 and 8, which are exposed on the hCYPA surface, are shown in green

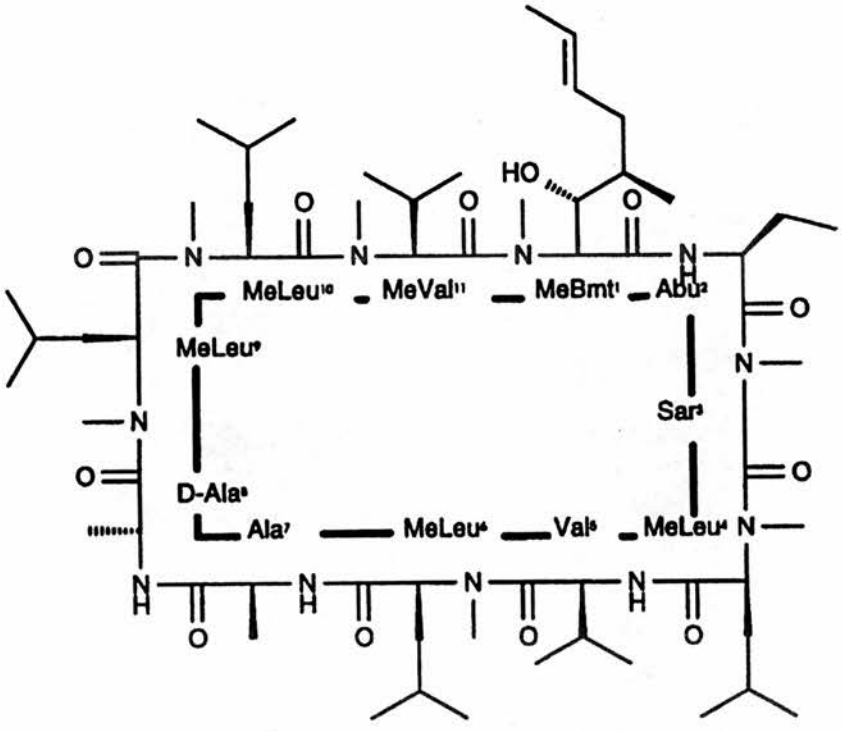


Figure 1-9. Chemical structure of CsA. Abbreviations: MeBmt for (4R)-4-[(E)-2-butenyl]-4-N-dimethyl-L-threonine; Abu for L-α-aminobutyric acid; Sar for sarcosine; MeLeu for N-methylleucine; MeVal for N-methylvaline

Moreover, there is one *cis* peptide bond between MeLeu9 and MeLeu10. However, when bound to cyclophilin, CsA adopts a loose folded conformation without any intramolecular hydrogen bonds. Additionally, all peptide bonds are present in the *trans* conformation. In contrast to CsA, cyclophilin does not undergo dramatic structural change upon ligand binding (Mikol et al., 1993).

(b) *Structures of hCYPA complexed with 11 CsA derivatives*:

A total of 11 structures of CsA and CsA derivatives bound to hCYPA have been refined at resolutions between 2.2Å and 1.76Å with R-factors varying from 15% to 19% (Kallen *et al.*, 1998). The formulae of 11 CsA derivatives and their biological activities are summarised in Table 1-F.

Comparison of these 11 structures (Kallen *et al.*, 1998) shows that the conformation of the cyclophilin is very similar. The maximum rms deviation among all structures is only 0.2Å. Apart from the flexible N- and C-terminal regions, the only region that reveals dramatic difference is the 70s loop (residues 67 to 76) (see Figure 1-7). The movement of this loop results from the binding of the CsA analogues to the protein. Comparison of CsA derivatives with CsA shows that the large difference occurs in residues 4 to 8 of CsA. This region (residues 4 to 8) is suggested to have specific interaction with calcineurin as mentioned before. In addition, the intermolecular hydrogen bond and hydrophobic interactions between cyclophilin and different CsA derivatives are conserved. This indicates that the binding site of hCYPA has a very conserved conformation even when different ligands bind.

As seen in Table 1-F, chemical modification of residues 1, 2, and 3 changes the ability of CsA to bind cyclophilin without affecting the immunosuppressive activity dramatically. Conversely modification of residue 4 decreases the immunosuppressive activity dramatically without reducing the binding affinity. This result is in agreement with the structural study where residues 1, 2, and 3 are responsible of the binding to cyclophilin while residue 4 affects the interactions with calcineurin. Residue 4 of CsA is a key site in the recognition of CYPA/CsA complex by calcineurin. A slight modification of the side chain of residue 4 is enough to essentially abolish the calcineurin inhibitory.

Table 1-F. Formulae of CsA derivatives and their biological activities

CS Deriv	R_1	R_2	R_3	R_4	CypA[a)] (%)	IL2[a)] (%)
CsA			H		100	100
116450		*	*	*	3.8	23.8
33804	*		*	*	16.6	32.2
27402	*		*	*	71.42	100
224698	*		*	*	13.8	14.3
209313	*	*		*	303	41.6
209650	*			*	144.9	41.6
209217	*			*	416.6	100
209825				*	188.6	55.5
211810	*	*	*		100	0.9
211811	*	*	*		166.6	< 0.06

Key:
(1) CsA=Abu2-CS; 116450=MeBm2t-CS; 33804=Val2-CS 27402=Thr2-CS; 224698=(5-hydroxy)Nva2-CS; 209313= D-MeSer3-CS; 209650=Val2-d-MeAla3-CS; 209217=Val2-D-(2-S-methyl)Sar3-CS; 209825=(6,7-dihydro)MeBmt-1-Val2 -D-(2-S-methyl)Sar3-CS; 211810=(4-hydroxy)MetLeu4-CS; 211811=MeIle4-CS
(2) The CYPA value in column 6 gives a measure of the strength of the binding the derivative relative to CsA as determined using an ELISA assay (Quesnizus *et al.*, 1988)
(3) The number given is column 7 is the measure of the effect of supressing the production of interleukin-2 relative to CsA in a whole cell assay (Fliri *et al.*, 1993; Bollinger *et al.*, 1990)
(4) [a)] : IC^{50} (derivative)/IC^{50} CsA

The ability to decouple the cyclophilin inhibitory property from its immunosuppressive activity by such modification is of potential value in developing non-immunosuppressant derivatives which prevent cyclophilin being incorporated into the HIV protein coat. It has been shown that MeIle4-CS (211811) is a potentially useful HIV inhibitor which blocks viral replication (Rosenwirth *et al.*, 1994).

1.2.6 PPIase (peptidyl-prolyl *cis-trans* isomerase) activity

The torsion angle of a peptide bond can be either *cis* (0°) or *trans* (180°). *Cis* peptides are rarely found in native structures of proteins. The rotational energy barrier of the peptide bond from *trans* to *cis* form has been calculated to be 21kcal/mol (Jorgensen and Gao,1988), except the Xaa-Pro peptide bond. Xaa-Pro has a higher probability (10% ~30%) of forming a *cis* peptide bond while other peptide bonds have a probability of 10^{-3}. The higher probability of Xaa-Pro in *cis* peptide bond is due to a lower rotational energy barrier of 17kacl/mol (Fischer *et al.*, 1994). Cyclophilin is known as a PPIase since their enzymatic activity can reduce the rotational energy barrier and speed up the *cis-trans* isomerization. The PPIase activity of cyclophilin plays an important role in the cell since it could accelerate the refolding of other proteins. Some studies have shown the *cis-trans* isomerization of the Xaa-Pro peptide bond is the rate-limiting step in the refolding process (Brandts *et al.*, 1975; Lang *et al*, 1987). The rate of refolding of RNAase was accelerated in the presence of cyclophilin at a concentration as low as 0.1μM (Schonbrunner *et al.*, 1991). When the concentration of cyclophilin increased to 30μM, which is the approximate concentration of some cyclophilins in *vivo*, RNAase was instantly refolded (Harding *et al.*, 1986). Addition of PPIase has also been shown to accelerate the refolding process on human carbonic anhydrase II (Fransson *et al.*, 1992). The half-time for recovery of enzymatic active of carbonic anhydrase II on refolding was reduced to 4min in the presence of PPIase compared to 9min in the absence of PPIase. Moreover, studies have been reported which illustrate the involvement of cyclophilins in the folding of several proteins: collagen or procollagen (Bachinger, 1987; Davis *et al.*, 1989; Steinmann *et al.*, 1991), chymotrypsin inhibitor 2 (Jackson and Fersht, 1991) and ribonuclease (Lang *et al.*, 1987; Kiefhaber *et al.*, 1990a,b,c).

The *cis-trans* isomerization can be considered as a two-step reaction (Ke *et al.*, 1993a). The first step is the rotation of the C-N peptide bond. The second step is the release of the N- or C-terminal residues of peptides or substrates from the PPIase. Since the C-N peptide bond has partial double bond character which results in the restriction of the rotation of C-N bond, the *cis-trans* isomerization needs to overcome the energy barrier. Therefore, any factors which can weaken the double bond character of the C-N peptide bond might be expected to accelerate the rate of isomerization.

Several mechanisms have been proposed for the *cis-trans* isomerization catalysed by cyclophilins. They will be discussed as follows:

(a) *Catalysis by formation of a covalent tetrahedral intermediate* (Fischer *et al.*, 1989a,b):

In this mechanism, it was suggested that a tetrahedral intermediate, hemithioorthoamide, was formed by the nucleophilic addition of a cysteine or nucleophilic residue to the carbonyl carbon of the Xaa-Pro peptide bond. The formation of the covalent intermediate is then followed by rotation about the C-N peptide bond and finally, collapse of the tetrahedral intermediate to release the isomerized Xaa-Pro peptide. However, the structural study revealed no cysteine or nucleophilic residue close to the carbonyl carbon (Ke *et al.*, 1991). Additionally, a site-directed mutagenesis study also disagreed with this mechanism as the mutation of cysteine to alanine has no effect on the *cis-trans* isomerase activity (Liu *et al.*, 1990).

(b) *Catalysis by distortion* (Harrison and Stein, 1990a, b):

In this mechanism, the C-N peptide bond is rotated and this distorted intermediate is stabilised by the noncovalent interaction with cyclophilin. The energy that is required to distort the peptide bond out of planarity and therefore destroy the resonance stabilisation of the amide linkage, is provided by the binding interaction between cyclophilin and substrate in the transition state. This mechanism is supported by the kinetic isotropic study and the solvent deuterium isotropic effect (Harrison and Stein, 1990a,b). However, from structural studies, all proline-containing peptides and substrates bind to CYPA in the *cis* form, except the HIV-1 capsid protein (Gamble *et*

al., 1996). Therefore, the distortion binding mechanism is not proven by the structures of cyclophilin complexed with a variety of proline-containing peptides as no distorted peptides in the transition state bind to cyclophilin.

(c) *Protonation of the nitrogen of C-N peptide bond* (Kofron *et al.*, 1991):

The mechanism proposes the side chains of serine, threonine or tyrosine protonate or form a hydrogen bond to the lone pair of nitrogen atom of C-N peptide bond. It has been proven by molecular orbital calculation that the protonation reaction of amide nitrogen atom can reduce the energy barrier of the *cis-trans* isomerization as it can deconjugate the C=O-N peptide bond and thus weaken the double bond character of the C-N bond (Armbruster and Pullman, 1974). However, this mechanism is not in agreement with structural results. In the structure of Ala-Pro binding to hCYPA, there are no proton-donor groups such as serine, threonine or tyrosine close to the amide nitrogen. Therefore, this mechanism is questionable and requires further investigation.

(d) *Catalysis by desolvation* (Radzicka *et al.*, 1992; Eberhardt *et al.*, 1992):

This mechanism suggests the desolvation of the proline-containing peptides or the substrates by binding to the PPIase may be the first step in the *cis-trans* isomerization. Further actions, such as hydrophilic interactions between the PPIase and substrates are required to complete the *cis-trans* isomerization. This mechanism is supported by the results that the rate of *cis-trans* isomerization was significantly accelerated in nonpolar solvents (Radzicka *et al.*, 1992; Eberhardt *et al.*, 1992). In addition, this mechanism is also supported by the structural analysis of cyclophilin complexed with a variety of substrates as the proline residue of the substrate makes strong hydrophobic interactions with the hydrophobic pocket of the cyclophilin.

(e) *Solvent-assisted mechanism* (Ke *et al.*, 1993a):

This mechanism is based on the structure of Ala-Pro binding to hCYPA. Ala-Pro is firstly desolvated in the ground state by binding to the hydrophobic pocket of the active site. Then, the C-N peptide bond of Ala-Pro is oscillated by thermodynamic fluctuation or thermal movement. Once this C-N peptide bond is rotated to an angle between -50° to -90°, a water molecule, which is conserved in the unliganded structure of hCYPA and complexed structure of dipeptide binding to hCYPA, can

form a hydrogen bond with the carbonyl oxygen atom of Ala-Pro and consequently stabilise the intermediate in the transition state. This solvent-assisted mechanism is supported by the structure of dipeptide binding to hCYPA. However, this mechanism is inconsistent with the structure of cyclophilin complexed with sAAPFna (Zhao *et al.*, 1996a) as this water is displaced by the oxygen of the succinyl group. The absence of the water molecule in the tetrapeptide complexed structure raises the question whether the dipeptide and tetrapeptide have different *cis-trans* isomerization mechanisms or whether solvent-assistant mechanism is wrong. Further experiments using NMR or kinetic method are required in order to answer this question.

1.2.7 Functions

Many functions have been attributed to the cyclophilins

(a)Immunophilin:

The most well-known function of cyclophilin is its immunosuppressive activity, which has been facilitated to prevent organ rejection after transplant operation. In order to initiate the immunosuppressive activity, CsA is required to bind to cyclophilin. Neither cyclophilin nor CsA alone can effect this activity. The mechanism of the immunosuppressive activity is complicated. Firstly, cyclophilin forms a tight complex with CsA and this complex will then bind to calcineurin and inhibit its phosphatase activity (Liu *et al.*, 1991, 1992; Clipstone et al., 1994). Calcineurin is a phosphatase and plays an essential role in the regulation of T lymphocyte activation and proliferation. Inhibition of calcineurin will inactive the dephosphorylation of a lymphocyte-specific transcription factor, NF-ATp, and therefore prevent the movement of NF-ATp from the cytoplasm to the nucleus where it binds to the promoter of the interleukin-2 gene and stimulates expression of the gene in the T-cell activation pathway (Loh *et al.*, 1996; Luo *et al.,* 1996).

(b)The PPIase activity:

The other important function of cyclophilin is its PPIase activity. Cyclophilin can catalyse the *cis-trans* isomerization of peptidyl-prolyl amide bonds and consequently accelerate the rate of refolding of certain proteins in the cell. The characteristic of

PPIase activity and its influence on several proteins *in vitro* have been discussed in section 1.2.6.

(c) *The incorporation of cyclophilin A into HIV-1 virions*:

Recently, it has been reported cyclophilin A forms a tight complex with HIV-1 capsid protein and the binding of cyclophilin to HIV-1 capsid protein is required for the infectious activity of HIV-1 virion (Franke *et al.*, 1994; Thali *et al.*, 1994). Therefore, any factors that disrupt the interaction between capsid protein and cyclophilin will then inhibit the replication of HIV-1. For example, CsA, the high affinity substrate of cyclophilin, significantly inhibits the incorporation of cyclophilin into HIV-1 virions and thus reduces the HIV-1 infectivity. As can be expected, if the inhibitory property of CsA can be decoupled from its immunosuppressive activity by chemical modification, this modified non-immunosuppressive CsA analogue could block HIV-1 virion replication without initiating the immunosuppressive reaction. Recently, it has been reported that MeIle4-CS, a non-immunosuppressive CsA analogue, is a potentially useful HIV virion inhibitor (Rosenwirth *et al.*, 1994). Furthermore, crystallographic and genetic studies propose that the CYPA plays a role as a chaperone or essential functional component instead of a *cis-trans* isomerase in HIV-1 infectivity as discussed in section 1.2.4.

(d) *The application in the field of parasitology*:

Cyclophilins have an important role in the field of parasitology as a large and increasing number of conserved and divergent cyclophilins have been purified and isolated in both protozoan, helminth and nematode species (Page *et al.*, 1995a,b; Bell *et al.*, 1996). Moreover, CsA, the well-known substrate of cyclophilin, has been shown to have anti-parasitic effects against a variety of parasites, such as, nematode species (Page *et al.*, 1995a,b). The mode of action of CsA or its analogues against parasites is still unknown. However, it has been proposed that the anti-parasitic properties of CsA are involved in producing structural damage to tegumental and gut surface of the parasites and is likely to be related to the inhibition of the folding of some proteins which are catalysed by PPIase under normal conditions (Page *et al.*, 1998).

(e) *Other important functions of some divergent cyclophilins*:

Some divergent cyclophilins have been shown to have the distinct functions, separated from the immunosuppressive function. The cyclophilin homology ninaA gene product from *Drosophila melanogaster* has been shown to form a tight complex *in vivo* with rhodopsin and this 26 kDa cyclophilin homologue is likely to be a transporter transferring rhodopsin from the rough endoplasmic reticulum through the cytoplasm to the cell surface. The ninaA gene is also probably to involve in the folding of an eye-specific rhodopsin class in *Drosophila melanogaster*. The inhibition of this folding has resulted in impaired vision (Colley *et al.*, 1991; Schmidt *et al.*, 1994). In addition, two cyclophilin-like proteins together with other proteins have been shown to chaperone the opsin molecule of the visual pigment of vertebrates (Ferreira *et al.*, 1996).

Moreover, hCYPB together with a heat shock protein (hsp47) has been reported to transport procollagen from endoplasmic reticulum to pre-golgi intermediate (Smith *et al.*, 1995).

Recently, it has been reported that cyclophilin 40 together with other proteins, forms the avian progesterone receptor (PR) (Johnson and Toft, 1994). This receptor is a large multiprotein complex containing CYP40, P23, two heat shock proteins (hsp70 and hsp90) and two FK506-binding proteins, FKBP54 and FKBP52. In the absence of hormone, PR has been shown to bind and respond to progestin.

1.3 Thesis rationale

The work described in this thesis involves the examination of the structure and ligand binding interactions of BLG and CYP3.

The determination of the structure of BLG in the lattice Y form (BLGY) has not progressed smoothly since it was first reported in 1979 at low resolution (Green *et al.*, 1979). An initial goal of this thesis was to collect a good quality data set at low temperature and to achieve a successful refinement of the BLGY structure (chapter 3).

The ligand binding properties of BLG are ill-understood at the molecular level although the protein has been shown to bind a variety of hydrophobic small molecules by solution studies. Therefore, one of the main goals of this thesis was to

provide a detailed picture of the binding site and consequently to understand the interaction between a ligand and BLG using x-ray crystallography (chapter 4). Similar to BLG, cyclophilin 3 (CYP3) is a small globular protein with molecular weight around 18000Da. Superficially, CYP3 shares a similar structural motif to BLG as both proteins contain a β-barrel formed by the eight antiparalled β-strands. However, the ligand binding characteristics and biological features of CYP3 are very different from that of BLG. A further aim of this thesis was to refine the CYP3 structure and to examine the ligand binding interactions of the protein by x-ray crystallography. Finally, by using the structure of CYP3 complexed with the dipeptide, Ala-Pro, as a model system, x-ray crystallography was applied to a dynamic study of ligand binding *"in crystallo"*. It has been shown that it is possible to obtain the dissociation constant by x-ray crystallography (chapter 5).

2. Materials and Methods

This chapter will describe the methods applied to BLG and CYP3. First, the materials used in the project (section 2.1) followed by the purification (section 2.2) and crystallisation conditions (section 2.3) will be reported. Next, the crystallographic data collection (section 2.4) and data processing (section 2.5) will be described. Finally, the theory and programs used in molecular replacement and refinement will be discussed in section 2.6 and 2.7, respectively.

2.1 Materials

- Bovine β-lactoglobulin variant B (Lot 13H7150) was purchased from Sigma Chemical Company.
- Phastgel Homogeneous 12.5 (Code 17-0623-01) was a 12.5% polyacylamide gel (PAGE gel) supplied by Pharmacia Biotech (Uppsala, Sweden)
- Phastsystem was supplied by Pharmacia. It consists of a separation and control unit and a development unit.
- The crystallisation screen I (HR-CS1) was supplied by Hampton research.
- Immersion oil (Type B) was supplied by Cargille Inc.
- Palmitic acid (Lot 105F00232) was 99% purity and supplied by Sigma Chemical Company
- Ammonium sulphate was ≥ 99% purity and supplied by Sigma Chemical Company
- HEPES (Lot 10077840W) was 99% purity and supplied by BDH
- MPEG5000 was supplied by Sigma Chemical Company
- Ala-Pro was supplied by Sigma Chemical Company
- Sodium citrate was supplied by Sigma Chemical Company

2.2 Purification

2.2.1 BLG

The BLG used in this work was obtained in two ways. One batch was purchased from Sigma. The other one was purified at the Hannah Research Institute.

2.2.1.1 Genetic typing of the Hannah Herd

Two genetic variants but three possibilities, BLG-A, BLG-B and BLG-AB, are found in the milk of Hannah herd. Therefore, it was necessary to type genetically the milk of several cows in order to identify individuals producing the pure A and pure B variants. Fresh milk was collected from the Hannah dairy in the morning. Analysis of all samples started within 3 hours of milking.

Genetic typing involved two stages. First, the whey protein was isolated from the whole milk. The second stage used "Ultrathin layer gel electrophoresis" to identify the variants. Ultrathin layer gel electrophoresis is very similar to the conventional gel electrophoresis but a smaller gel is used in the ultrathin layer gel electrophoresis. This method is faster than conventional gel electrophoresis and is supposed to give high resolution. The protocol used was as follows:

(a) Isolation of the whey protein;

First, fat was removed by skimming the milk. The milk was centrifuged at 40g for 25 mins and then cooled down in a refrigerator to solidify the fat layer. After 40 mins, a slightly yellow fat layer was evident at the top of the centrifuge tube. Skimmed milk was in the lower layer and some impurity was precipitated at the bottom of the tube. The fat layer was removed with a spatula. 50ml skimmed milk was taken out of tube with the volumetric pipette without disturbing the impurity at the bottom.

In the next step, casein was separated from whey protein by precipitation at the isoelectric point. 1M HCl was slowly added to 50ml skimmed milk to reduce the pH from 6.6 to 4.6, the isoelectric point of casein. Casein started to precipitate when the pH fell below 5.0. The cloudy solution was filtered through a 13mm polypropylene Millipore filter unit (0.45μm poresize). After filtering, the slightly cloudy solution containing whey protein was collected and stored in the refrigerator for further use.

(b) Identification of the variant by ultrathin layer gel electrophoresis

Buffer containing 11mM Tris and 53mM glycine at pH 8.6 was prepared. A pre-cast homogeneous 12.5% polyacylamide gel was soaked in approximately 100ml Tris/Glycine buffer for at least 1 hour before use. After soaking, the gel was removed and stood upright to drain on a tissue for around 10-15 mins. The gel was ready for use in the Phast system. The buffer strips to connect the gel to the electrodes were prepared according to the following protocol. 1g Agarose was dissolved in 50ml Tris/Glycine buffer using a hot plate and stirred with a thermometer until the temperature reached ~ 85°C. The solution was clear at this stage. Then, the clear solution was allowed to cool down to 58 °C and poured into the moulds using 2.4 ml per mould. The strips were covered with parafilm and stored in the refrigerator.

The Pharmacia Phast system was thermostated at 15°C before use and the base of the separation plate was cleaned with alcohol and wiped dry. Then distilled H_2O =~ 200µl was put in the centre of the base plate. The gel was placed carefully on the separation plate to avoid bubbles between the plate and the gel. After that, the buffer strips were carefully placed in the appropriate position. The electrode plate was then put down making sure that the electrodes were in contact with the buffer strips.

Meanwhile, the sample solutions were prepared according to the following protocol. Since the approximate quantity of BLG in milk was 3g BLG/ per litre milk, in order to load a protein concentration of approximately 500µg/ml, the whey protein was diluted 6 times before being applied to the gel. This was achieved by dilution of 20µl of whey protein with 110µl of the Tris/Glycine buffer and 10µl bromophenol blue solution.

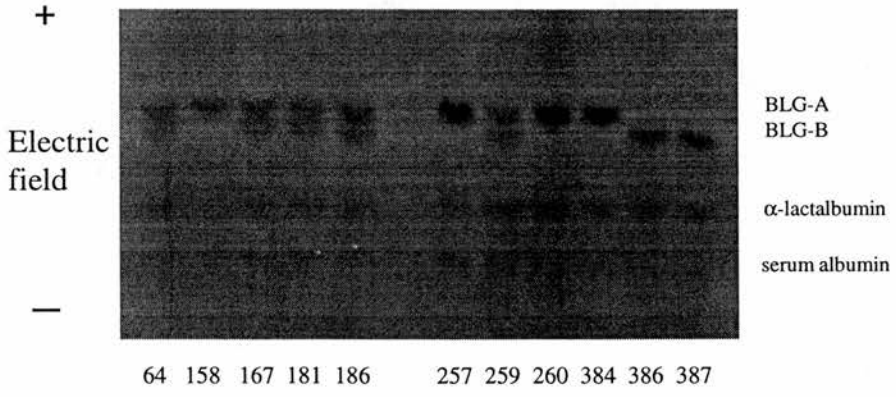
Sample solutions were then loaded into the gel according to the manufactory instructions to start gel separation procedure, which took around 10mins.The protein was stained with coomassie blue. Because variant A was more negative than B, it moved faster under the applied electric field which consequently separates these two variants.

The result of genetic typing of some of the Hannah herd is summarised in Table 2-A and a photograph of the gel is shown in Figure 2-1.

Table 2-A. The variants of BLG in milk from Hannah Herd

Cow number	Lactation number	Lactation date	Genetic variant
64	3rd	12/95	AB
158	2nd	8/8/95	A
167	2nd	29/8/95	AB
181	2nd	29/8/95	AB
186	2nd	2/8/95	AB
257	1st	9/95	A
259	1st	30/7/95	AB
260	1st	8/95	A
384	1st	10/95	A
386	1st	10/95	B
387	1st	11/95	B

Figure 2-1. Photograph of the gel used for genetic typing. The numbers represent the numbers of cows labelled in the Hannah herd.



2.2.1.2 Purification of BLG

The "salting out" method (Armstrong, 1967) was chosen to purify BLG. The purification procedure involves four stages:
the removal of fat, the removal of casein, fractionation of whey protein and the final purification of BLG. The milk of cow no.386 (variant B) and cow no.384 (variant A) were chosen to be purified. The protocol was as follows:

(a) First ammonium sulphate Precipitation:

Ammonium sulphate (264g/per litre of milk) was slowly added at room temperature over 40 mins. The solution was left to stir for 2 hours and then filtered through a Whatman 113V fluted filter paper. If the solution was still cloudy, it was then filtered through No. 42 filter paper overnight in the cold room to remove fat and some of the casein.

(b) Acid precipitation:

The faintly cloudy solution was taken from the cold room and brought to room temperature. 1M HCl (38 ml/per litre milk) was slowly added to reduce the pH from 7.04 to 3.5. The casein started to precipitate when the pH was below 5.0. The solution was left to stir for 1 hour and then centrifuged at 70g over 40 mins at 16°C using a Mistral 6000 centrifuge. After centrifuging, the supernatant was carefully poured off and filtered through No 42 filter paper. 1M NH_4OH was added to re-adjust the pH of supernatant to 6.0.

(c) Second Ammonium Sulphate Precipitation:

Ammonium sulphate (262g/per litre of the supernatant obtained from the previous step) was gradually added with stirring. The procedure was done very slowly in order to get a high yield. White BLG precipitated at this stage. The solution was left to stir overnight in the cold room to allow complete precipitation of BLG. Then, the milky solution was centrifuged at 80g in room temperature for 30 mins. After the supernatant was poured off carefully, the white precipitate was collected with a spatula and the tube was washed with a minimum amount of supernatant. The precipitate was mixed with a small amount (~2ml) of acetate buffer for dialysis.

(d) Dialysis:

Before dialysis, a suitable length of Visking tubing (18/32" diameter) was boiled for one hour in Na_2CO_3 in order to remove the plasticizer and any heavy metals. The precipitate of BLG was packed to the tubing and dialysed against 0.12M sodium acetate/0.04M acetic acid buffer, pH 5.2 overnight. The buffer was changed 4 times at 4 hourly intervals. Next morning, there was some black impurity, α-lactoferrin, precipitated at the bottom of tubing. The solution was centrifuged at room temperature for 30 mins by Sorvall centrifuge using the SS34 rotor. After removing α-lactoferrin, the supernatant was dialysed against phosphate buffer at pH 6.7 (0.02M Na_2HPO_4/ 0.1M NaCl).

(e) Gel Filtration:

A column of length 100cm and diameter 2.5cm was packed with Sephadex G-75, equilibrated in 0.02M Na_2HPO_4/ 0.1M NaCl and 0.02% azide, pH 6.7 over 72 hours. The buffer was de-gassed before use. Before loading into column, the concentration of protein in the dialysed sample was determined using ε_{1cm} =9.6 ml/mg*cm at 278nm by measuring its UV absorbance. Where the concentration was too high, the sample was diluted with phosphate buffer to ~ 30mg/ml. The flow rate was set to 0.4ml/min and the BLG eluted within approximately 270ml.

(f) Freeze-Dried the protein:

The fractions with high absorbance UV value was collected and dialysed against de-ionised H_2O for one day. Then the solution was freeze-dried for storage. The protein, which looked like white cotton after freeze-drying, was stored in the freezer until required. The purity of the final sample was checked by ultrathin layer gel electrophoresis. A photograph of the gel is shown in Figure 2-2

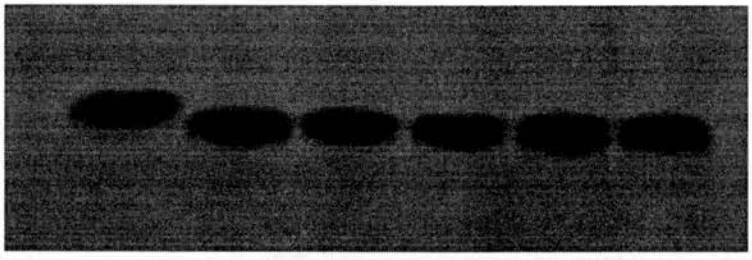
2.2.1.3 Result

(a) By weighting the BLG after freeze-dried, the yield obtained was as follows:

BLG-A: 326mg (326mg/1L milk)

BLG-B: 2500mg (1000mg/1Lmilk)

(b) Figure 2-2. Photograph of the gel used to check the purity of BLG after purification. The numbers represent the numbers of cows labelled in the Hannah herd.



2.2.2 BLG-palmitate complex

Bovine BLG-AB complexed with palmitate was prepared by Puyol in Spain. The procedures are included for completeness.

The procedure was according to Puyol *et al* (1994) and BLG-AB was delipidated by charcoal treatment at pH 3 as described by Chen *et al* (1967). Palmitic acid at a molar concentration ratio of 2:1 with respect to the protein dimer was dissolved in chloroform in a glass tube. After the organic solvent was evaporated under nitrogen, a solution of delipidated BLG, dissolved in 0.29 M NaCl, 2.5mM KH_2PO_4, 16mM K_2HPO_4 pH 7.4 was added and the mixture incubated overnight at 37 °C. The saturated protein solution was then dialysed against distilled water and freeze-dried. Analysis by gas chromatography showed that the complex had about 1 mol palmitate bound to 1 mol of dimeric protein.

2.2.3 CYP3

CYP3 was cloned by Dr. Page in Glasgow University and was purified by Miss Jacqueline Dornan in the laboratory in Edinburgh. The procedures are included for completeness.

2.2.3.1 Cloning of CYP3

CYP3 was cloned by Dr. Page (Page *et al.*, 1996) using the reverse transcriptase polymerase chain reaction (RT-PCR) method. The RT-PCR conditions were as follows: 10pmol of each primer (CYP3F/Nde CGCATATGAGCCGCTCAAAGGTCTTTTTTCG and CYP3R/EcoRI GCGAATTCTTATGCCTTGAGTTGTCCACAG), Buffer (45mM Tris pH 8.8; 11mM ammonium sulphate; 4.5mM $MgCl_2$; 6.7mM 2-ME; 4.4µM EDTA pH=8; 1mM of each; 113µg/ml non-acetylated BSA), 2µl of mixed stage cDNA and 1µl Taq polymerase. All of the above solution were mixed together and cycled 20 times at 94°C for 30secs (denaturation), 60°C for 1min (annealing) and 72°C for 1min (extension) in 100µl. Plasmid DNA was isolated (Plasmid Kit, Qiagen) and the insert was sequenced in both directions to check that no mutations had been generated by

the RT-PCR procedure. The plasmid DNA was then transformed into BL21 (DE3 pLysS) cells (Promega) for recombinant protein expression. After expression, the protein extract containing recombinant CYP3 was subject to purification.

2.2.3.2 Purification of CYP3

CYP3 was purified by Jacqueline Dornan in the laboratory in Edinburgh. First, protein extract was loaded into a SP Sepharose cation exchange column. Two buffers were used in sequence-- Buffer A: 50mM HEPES, 5mM EDTA, 5mM β-Me at pH6.8; Buffer B: Buffer A mixed with 0.5M NaCl, pH=6.8. Column fractions were tested by 15% SDS-PAGE gel. The fractions that contained CYP3 were collected and dialysed against buffer A overnight. After filtering through a 0.2μm filter, the solution was loaded into a Resource S ion exchange column. CYP3 was eluted within 75ml by the following buffer-- Buffer 1: 20mM HEPES at pH=6.8; Buffer 2: Buffer 1 mixed with 0.5M NaCl.

2.3 Crystallisation

2.3.1 Introduction

2.3.1.1 The theory of crystallisation

Crystallisation is a multiparametric process and involves three steps: nucleation, growth and cessation (Stura and Wilson, 1992).

First, crystallisation starts by nucleation, in which the macromolecule (protein) free in solution forms the first ordered aggregates. In order to achieve nucleation, the protein solution is brought into a supersaturated state, which is a metastable condition where more protein is in solution than is in thermodynamic equilibrium. A supersaturated state is thermodynamically unstable and may return to equilibrium by forming a crystalline or amorphous phase. Once the nucleation has occurred, the crystallisation goes on to the second step, crystal growth. Crystal growth is a process where protein is converted from the solution phase to the crystalline phase. In this step, protein is crystallised by the slow precipitation of the protein onto an ordered crystal lattice. When the crystals are formed by the addition of protein molecules onto

the growing faces of the crystal lattice, the degree of supersaturation is decreased as well.

Cessation of crystal growth can occur for many reasons. The most obvious is due to accumulation of impurities in solution, which results in growth defects, poisoning of the faces or ageing of the macromolecules.

There are various parameters responsible for the solubility of a protein and consequently they will affect the crystallisation process. These parameters (Blundell and Johnson, 1976), such as precipitant type and concentration, pH, ionic strength, temperature and the presence of additives (e.g. organic solvent) which change the dielectric constant of the solution need to be explored to optimise the crystallisation conditions. Because a variety of parameters need to be considered, in order to reduce the number of experiments and the consumption of protein, a more practical approach using sparse matrix screens is usually carried out (Jancarik and Kim, 1991). In the sparse matrix screening method, the crystallisation conditions have been chosen based on a data base of successful protein crystallisation. These conditions include reagent combinations as well as the pH ranges which are most frequently reported in the literature. The concept of sparse matrix screening has become widely accepted and these crystallisation conditions are now commercially available as kits (Crystal Screen, Hampton Research).

Once success in a sparse matrix screen has been achieved, a grid screen should be performed to optimise the crystallisation conditions in order to obtain better crystals. Grid screens generally screen a single precipitant concentration versus a range of pH based on the information obtained from the sparse matrix screen. Grid screens are useful to fill in pH and reagent gaps missed with sparse matrix screens.

Four methods (Weber, 1997) are commonly applied to the crystallisation of macromolecules: vapour diffusion, free interface diffusion, batch and dialysis methods. The vapour diffusion method is used for the work described in this thesis. The vapour diffusion technique is the most popular method for the crystallisation of a protein as it is easy to set up, cheap and most important, it consumes little protein. Therefore, vapour diffusion is extensively used in screening for the suitable crystallisation conditions. Typically, the drop, either a hanging or sitting drop,

containing the 1:1 mixture of protein and precipitant is suspended and sealed over a well or reservoir solution, which contains the precipitant at the target concentration. The principle of vapour diffusion is straightforward. Since the drop contains a lower precipitant concentration than the reservoir, in order to reach equilibrium, water moves from the drop to the reservoir. As water leaves the drop, both the protein and precipitant concentration increase to achieve the supersaturated state and therefore lead to crystallisation.

2.3.1.2 The theory of soaking and co-crystallisation

Inhibitors, activators, substrates, or heavy atoms can be introduced into a crystal to bind to the protein in two ways (Stura and Chen, 1992). One is to soak native crystals of protein in the mother liquor containing ligand or reactant solution. Because a protein crystal has a high solvent content (between 27-95%), apart from a small amount of the solvent which is tightly bound with the protein, most solvent is weakly bound and can be replaced by ligand. The maximum size of the ligand that may be diffused into the crystal depends on the size and environment of the channels within the crystal. For most small molecules, these channels are large enough to allow their diffusion to any part of the surface of the protein except for the regions involved in crystal contacts. However, if the ligand is large or hydrophobic, diffusion into the crystal becomes inefficient. In these cases, co-crystallising ligand with protein is an alternative way to introduce the ligand into the crystal. Compared to the soaking method, co-crystallisation sometimes can produce crystals which are totally non-isomorphous with the native crystal, e.g. different space group, and this makes the structure determination more complicated.

2.3.2 BLG

2.3.2.1 Crystallisation of BLGY

Initial trials of crystallising Lattice Y used ammonium sulphate according to the existing recipes (Brownlow *et al.*, 1997, Hambling, 1990). However, they failed to produce good crystals suitable for X-ray diffraction when either the batch method or the vapour diffusion method were tried. Only amorphous precipitate appeared.

However, in some drops, microcrystals were obtained. After optimising the crystallisation conditions by varying protein concentration, pH, salt concentration and temperature, the best crystals obtained only diffracted to 3.5Å. Therefore, the crystal screen kit (Hampton research, HR-CS1) was used to screen for alternative crystallisation conditions. Six lattice Y crystals with a size around 0.5mm were obtained in reagent No.38, whose composition was 1.4M sodium citrate dihydrate, 0.1M HEPES, pH 7.5. The crystals diffracted to a resolution better than 3Å. Then, the crystallisation conditions were refined by varying the sodium citrate dihydrate concentration, pH, protein concentration. After several trials, good crystals that diffracted to 2Å were obtained by the following procedure:

Bovine BLG-B was dissolved in distilled H_2O to a concentration of 40mg/ml and crystallised at 16°C by the sitting drop method using 1.34~1.40M sodium citrate as precipitant and 0.1M HEPES adjusted with 1M NaOH to pH 7.5 as buffer. 0.02% azide was also added to prevent microorganism growth. Distilled H_2O was used through the whole experiment. A microbridge was placed in 1.0 ml of precipitant solution, the reservoir, in a 24-well Linbro tissue culture plate. For a sitting drop, 4µl protein solution was added to 12µl reservoir solution. After the well was sealed with a glass cover-slip and immersion oil, the whole tray was left in the 16°C crystal growth room. After 4~5 days, lattice Y crystals with size 1.0x0.8x0.9 mm, had grown from the clear drops (see Figure 2-3, top left).

BLG-A and BLG-AB were also crystallised using the same conditions as above. However, the quality of crystals, as judged both by the size and diffraction pattern, was not as good as those of the BLG-B variant. The BLG purified at the Hannah Institute herd was tried as well. Crystals with dimensions around 1mm were obtained from the same conditions. However, compared to the chunky crystals from the Sigma protein, the crystals were long and thin, which resulted in the elliptical spot shape in the diffraction pattern. Furthermore, the overall quality of the data was worse than that for Sigma BLG. The reason why the more pure protein from the Hannah herd produced worse crystals than the protein from Sigma is unknown. One possibility is that some low level impurity maybe plays the role of an additive to improve the crystal quality.

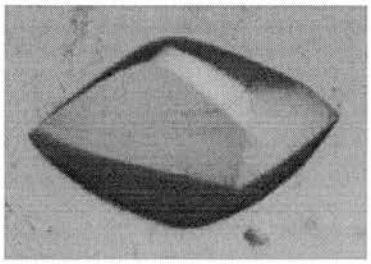


Figure 2-3. Photograph of a crystal of BLGY

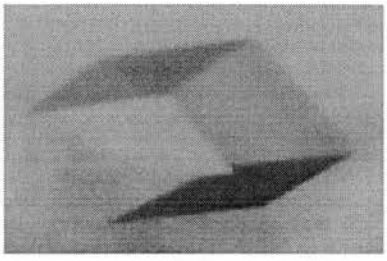


Figure 2-3. Photograph of a crystal of BLGZ co-crystallised with palmitate

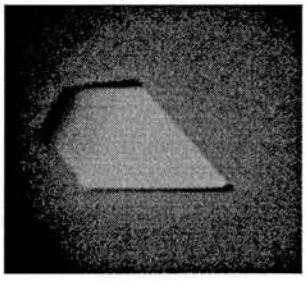


Figure 2-3. Photograph of a crystal of CYP3

2.3.2.2 Co-crystallisation of BLG with palmitate

Once native crystals were available, the soaking experiments were carried out to introduce ligands into the crystals. However, crystals tended to crack easily in the soaking solution, which contained 5%-10% organic solvent (for example, ethanol) in the mother liquor to help dissolve the hydrophobic ligand. Apparently, the additional organic solvent disturbs the interaction between molecules at the crystal contact in the lattice which results in cracking of the crystal. If the content of organic solvent was decreased below 5%, the crystals sometimes survived. However, this reduced the amount of palmitate available to diffuse into the crystal to bind to the protein since there was no promising electron density observable in the difference maps. Therefore, soaking experiments were abandoned and the alternative method, co-crystallisation, was used to try to introduce the palmitate into crystals.

BLG-B crystallisations were set up as described in ***2.3.2.1***, but 0.4µl 100mM palmitic acid in 100% ethanol was added to the drop to give a molar ratio of 10 palmitic acids per protein dimer, and mixed by pipette before the well was sealed with the glass cover-slip. Since the palmitic acid was super-saturated with respect to the aqueous phase, a white precipitate appeared in the drop. After 4-5 days, the white precipitate had disappeared and trigonal lattice Z crystals, with size 0.4x0.3x0.5 mm, grew from the clear drops (see Figure 2-3, top right).

The pH of the crystals was assumed to be 7.5, although it was not directly measured. It is interesting to notice that, in a control experiment, with the same conditions but no added ligand solution, lattice Y rather than lattice Z crystals appeared in the drop.

2.3.2.3 Crystallisation of BLG complexed with palmitate

The freeze-dried complex prepared by Puyol (refer to **2.2.2**) was dissolved in distilled H_2O to a concentration of 40mg/ml and crystallised as a sitting drop (4µl protein solution + 12µl well solution) over a well solution of 1.4M sodium citrate, 0.1M HEPES at pH 7.5 and 0.02% azide. Crystals, also of the trigonal lattice Z form, grew

in a few days and appeared identical to the crystals from the co-crystallisation preparation.

2.3.3 CYP3

Crystallisation and ligand soaking of CYP3 were done by Miss Jacqueline Dornan in the laboratory in Edinburgh. The methods are included for completeness.

2.3.3.1 Crystallisation of CYP3

CYP3 purified by the procedure outlined in section **2.2.3** was crystallised at 16°C by the hanging drop vapour diffusion method. The well solution contains 29~31% MPEG5000 as precipitant and 0.1M sodium citrate adjusted by citrate acid to pH 5.6. For a hanging drop, 2μl 7.0mg/ml CYP3 in 20mMHEPES/0.5MNaCl buffer was added to 2μl well solution.

Tetragonal bipyramidal crystals of size 0.2x0.3x0.3 mm appeared after 1~2 weeks (see Figure 2-3).

2.3.3.2 Soaking of CYP3 with Ala-Pro

The conditions of the native CYP3 crystals soaked in solutions with different concentrations of Ala-Pro are summarised in Table 2-B.

Table 2-B. The conditions of the native CYP3 crystals soaked in solutions with different concentrations of Ala-Pro

Ala-Pro Concentration(mM)	Soaking Vol. (μl)	Soaking time
1.2	10	27 hours
6	10	20 hours
18	10	20 hours
36	10	18 hours
60	10	20 hours
120	10	17 hours

2.4 Data Collection

The goal of data collection is to collect all unique reflections for a specific space group. Additionally, because crystals could decay in the x-ray beam due to reactive

free radicals generated by the radiation, protein crystallographers have to use an efficient data collection strategy to collect data as rapidly and precisely as possible. It is also highly advantageous to collect as many as symmetry equivalents as are accessible with in the limits of the experimental set up.

2.4.1 Preparing a crystal for data collection

Traditionally, crystals are mounted in a capillary with mother liquor to keep them from drying out during the x-ray experiment. However, crystals usually decay quickly at room temperature before a complete data set can be collected. This problem has been overcome recently by suddenly cooling the crystal in liquid nitrogen to a temperature around 100K, and thereafter keeping the crystal in the stream of gaseous liquid nitrogen during data collection.

This flash freezing method has been shown to increase crystal lifetime significantly, in some case even making crystals survive indefinitely (Dewan and Tilton, 1987). Moreover, the advantage of this method is that it is rather easy and there is very little mounting material and surrounding mother liquid to cause a significant radiation background. Recently, it has been found some protein crystals can be frozen first and unfrozen later for several rounds to collect more than one type of data set from the same crystal. Surprisingly, the mosaicity of the crystal decreased during this procedure (Yeh and Hol, 1998).

A suitable cryoprotectant is essential for the success of the flash freezing method. The function of the cryoprotectant is to prevent the crystal cracking when the disordered water and solvent in the crystal freeze as an amorphous glass. Some chemicals, such as ethylene glycol, glycerol or PEG400, at concentrations between 15-25% in mother liquor are frequently used as cryoprotectants. There are several ways to introduce the crystal into the cryoprotectant: direct transfer into the final concentration of cryoprotectant for a few seconds, sequential transfer into increasing concentrations of cryoprotectant and co-crystallisation with cryoprotectant (Garman, 1996).

The flash freezing procedures for BLG and CYP3 were as follows:

(a) BLG:

A crystal about 0.8mm in length was mounted in a 1.0mm cryoloop (Hampton Research, Inc.), dipped briefly (2-3 seconds) in immersion oil (TypeB, Cargille) and frozen by plunging into liquid N_2. The frozen crystal was then transferred to a magnetic goniometer head in a stream of N_2 at 100K (Cryostream, Oxford Cryosystems).

(b) CYP3 :

The composition of cryoprotectant required for CYP3 was: 0.2ml glycerol, 0.7ml mPEG5000, 0.1ml 0.1M sodium citrate/citric acid (pH 5.6). A crystal about 0.3mm in length was mounted in a 0.2~0.3mm cryoloop (Hampton Research, Inc.), dipped in cryoprotectant for a few seconds and frozen by plunging into liquid N_2. The frozen crystal was then transferred to a magnetic goniometer head in a stream of N_2 at 100K as for BLG.

2.4.2 Optimising data collection parameters

Although the quality of crystal is the most important factor that determines the quality of data, other factors also affect the success of data collection. It is impossible to fully satisfy all of the factors at the same time. Thus, any data set is usually the result of compromise.

These factors will be discussed below:

(a) Detector distance:

The crystal-to-detector distance is the first parameter to decide upon since it determines the spot separation and hence the maximum resolution that can be collected.

$$\text{Spot separation} = 1/L * D * \lambda$$

where L: the length of one of unit cell axes, normally the longest
D: detector distance
λ : wavelength

Therefore, the distance between spots depends linearly on the detector distance. The further the detector distance, the better the separation. However, if the detector distance is too far, the reflections can be beyond the edge of the detector and consequently the high resolution data will be missed out. A rule of thumb is to move

the detector back until the highest resolution reflections just reach the edge of the detector. This reduces radiation background and uses the detector surface most efficiently.

(b) Oscillation width per frame:

The permissible oscillation width per frame without spot overlap depends only on the maximum resolution and cell dimension parallel to the beam direction (Dauter, 1997).

$$\Delta\phi = (180/\pi)\ (d/a)$$

where d: the resolution limit
a: cell dimension parallel to the beam direction

Because crystals often have different cell dimensions in different directions, the maximum allowed rotation range may also change when the crystal rotates from image to image. However, the cell dimension along the spindle axis will never lie along the beam since the crystal is rotated around this axis. Thus, if a crystal has one cell dimension much longer than the other two, it is best to align this longest axis along the goniometer spindle axis.

(c) The required oscillation range:

How many degrees need to be collected in order to get a complete data set depends upon the crystal symmetry and the choice of the proper rotation start and end positions. The required oscillation range varies from 30° to 180° (Dauter, 1997) but assumes that the crystal is perfectly aligned. Usually, because the crystal is misaligned in practice, the required oscillation range is larger.

(d) Slit size:

The bigger the slit size, the stronger is the x-ray beam that will radiate the crystal and consequently, the spot intensity will be stronger. However, the background radiation also increases. Also, the spot size will be bigger with a wider slit and this may result in spot overlap. Thus, ideally, the slit size should be set equal to or smaller than the crystal size.

(e) Exposure time:

The longer the crystal is irradiated in the X-ray beam, the stronger reflections are recorded in the detector. Therefore, if the high resolution reflections are too weak, it

is sometimes useful to increase the exposure time in order to get stronger reflections. However, if the exposure time is too long, the low resolution reflections will start to overload. Thus, how to choose a reasonable exposure time is a result of compromise between the collection of low and high resolution data.

The other alternative way is to decrease the oscillation width per frame with the same exposure time. With the Mar detector, it is sensible not to have a maximum pixel intensity on any frame greater than 45000 although the intensity up to 105000 is processible (DENZO default number).

2.4.3 Summary of the optimised parameters for data collection

All diffraction data with the exception of the data from the native BLGY crystal were collected on a 300mm MarResearch imaging plate system mounted on an ENRAF Nonius FR571 rotating anode generator operating at 40kV, 80mA and producing Cu-Kα radiation from a graphite crystal monochromator. The pixel size of the Mar detector is 150μm x 150μm. The data of native BLGY was collected on a 180mm MarResearch imaging plate using the synchrotron source in Daresbusy 7.2 station.

Table 2-C. The summary of data collection parameters for BLG

Ligand	Distance (mm)	Slit size (mm)	Oscillation width (°)	Exposure time (sec)	Oscillation range (°)
Native*	83.9	0.2	1.4	120	0-182
Palmitate[1]	180	0.35	1.5	1500	0-132
Palmitate[2]	160	0.30	1.5	1200	0-67.5

native*: data was collected in Daresbusy using small mar detector
palmitate[1]: data collection of BLG co-crystallised with palmitate (first preparation, section 2.3.2.2)
palmitate[2]: data collection of BLG complexed with palmitate (second preparation, section 2.3.2.3)

Table 2-D. The summary of data collection parameters for CYP3

Ligand	Distance (mm)	Slit size (mm)	Oscillation width (°)	Exposure time (sec)	Oscillation range (°)
Native	120	0.3	1.2	1200	0-48
1.2mM Ala-Pro	145	0.3	1.2	1200	0-79.2
6mM Ala-Pro	135	0.3	1.2	1200	0-72
18mM Ala-Pro	130	0.3	1.2	1200	0-52.8
36mM Ala-Pro	130	0.3	0.9	1800	0-63.9
60mM Ala-Pro	140	0.3	1.25	1200	0-76.8
120mM Ala-Pro	135	0.3	1.2	1200	0-108.75

2.5 Data Processing

Data processing is a procedure to analyses, integrate, scale and merge the diffraction data obtained from data collection. The aim of this procedure is to index the recorded reflections, properly integrate the intensity and assess the quality of data set statistically.

2.5.1 Data processing protocol

Program DENZO (Otwinowski and Minor, 1997) was used to process the data. The program consists of the three subprograms: DENZO, XDISPLAYF and SCALEPACK. Subprogram DENZO separates the background from the reflection and writes these integrated reflections into a file containing hkl indices, intensity and estimates of the errors. XDISPLAYF is a program to display the diffraction image in order to monitor the data processing procedure. SCALEPACK is to scale the integrated reflections from DENZO and merge the symmetry-related reflections to a symmetry-unique set of reflections.

The protocol is as follows:

(a) *Autoindexing the diffraction pattern:*

The first step is to search for a number of well separated and medium intensity reflections to be used by autoindexing. Then, autoindexing is performed to search for all possible indices of all reflections picked from the peak search step by fast Fourier Transform (FFT) method. After the search for all possible indices is complete, the three linearly independent vectors with minimal unit cell volume are obtained. These three linearly independent vectors are able to index all of the observed reflections and form a basis for the determination of the true unit cell. The output of autoindexing in DENZO is the lattice and unit cell distortion table, and the crystal orientation parameters. The lattice and unit cell distortion table show the 14 Bravais lattices in decreasing order of symmetry, with primitive cubic being the highest symmetry and primitive triclinic being the lowest symmetry. Along with each lattice is a percentage value which represents the distortion of the unit cell parameters in order to fit the lattice. From the table, the lattice of highest symmetry with minimal distortion is chosen from among 14 Bravais lattices.

(b) Optimisation of spot size and mosaicity:

The spot size and shape were chosen to fit the reflection as closely as possible. It could be slightly bigger than the actual reflection size, but must not overlap the neighbouring reflections. The background size was chosen 0.1mm larger than the spot size.

Mosaicity is optimised to give more or less the same number of predicted reflections as are observed reflections. However, it is better to give slightly higher mosaicity to be sure of picking all the spots.

(c) Refinement of the crystal orientation and detector parameters:

Because the approximate orientation of the crystal and initial detector parameters given from the autoindexing step were imprecise, the crystal and detector parameters had to be refined in order to get an accurate prediction of the spot position. Minimisation of the deviation of the reflections from their predicted positions was essential for obtaining precise integrated reflection intensities. This least-squares refinement is applied in DENZO and the χ^2 value, a measure of the difference between predicted and recorded reflections, should be near 1.0 if the refinement has converged. However, a χ^2 value of 2 or even 3 is still acceptable. In the early stages of refinement, χ^2 values will be high, but when the orientation of the crystal, the detector parameters and the unit cell parameters start to be refined, the χ^2 values should decrease gradually and end up below 2. The predicted decrease is a measure of the convergence of the refinement. When the predicted decrease is near zero, the refinement has converged and further refinement will have no effect.

(d) Integrating the reflection intensity:

A profile fitting method was applied in DENZO to integrate the intensity. This method makes the assumption that the profile of the reflections is independent of the intensity, which means the strong and weak reflections have the same profile. Therefore, the strong reflections can be used to determine the profile and then this profile is applied to the weak reflections. Once the profile was derived from the strong reflections, this profile was fitted to each reflection as closely as possible and the area under the profile was counted as intensity. The rest of the area outside the

profile was then considered as background. The background must be subtracted from the intensity to get the correct measurement.

The profiles of the observed reflections vary over in the diffraction pattern as the profiles of the reflections at the centre of the diffraction pattern are more circular than that at the edges because the incident beam is more oblique at the edge of the detector. Thus, several profiles were usually used in the different parts of one particular image. In DENZO, each profile was derived from 10-50 strong reflections.

(e) Scaling and merging the integrated reflections:

Scaling data means to put all reflections on the same scale. Merging data means to find reflections with symmetry-related indices and transfer them to a symmetry-unique set of reflections. Therefore, the program SCALEPACK consists of three parts:

(1) calculate the single isotropic scale and temperature factor for each image or batch in order to scale all data to the same level

(2) reject the outliers and analyse for systematic errors

(3) average the symmetry-related reflections and merge them to a unique set data

2.5.2 Space group determination

The space group was determined by the following steps:

(1) The lattice type with highest symmetry and lowest distortion was selected from the distortion table in DENZO

(2) The data were scaled first in SCALEPACK by the lowest symmetry Laue group in the Bravais lattice selected from step (1). Then, the Laue group with higher symmetry was tested by the same procedure and the χ^2 was compared with the previous one from the lowest symmetry Laue group. The χ^2 from the SCALEPACK statistics output was the criterion to decide the Laue group. If χ^2 was significantly higher, the previous Laue group was the correct one. If χ^2 was almost the same, the next higher Laue group was tried. The procedure was repeated until the Laue group of highest symmetry with similar χ^2 was found.

(3) After the Laue group was determined, the systematic absences in the output log file were checked. Once the systematic absences were confirmed, the space group can subsequently be determined. However, some space groups have the same systematic absences, for example, the enantiomeric space group, $P3_121$ and $P3_221$. In these cases, scaling and merging of reflections can not distinguish between these enantiomeric pairs and hence these space groups have to be determined from the translation function in molecular replacement or by some other procedures.

2.5.3 Assessment of data quality

The quality of the data can be assessed from the SCALEPACK output statistics. There are two ways to justify the data quality:

(a) The ratio of intensity to the sigma of intensity, i.e. I/σ:

The sigma (estimated standard deviation) of an intensity is an estimate of how accurate the individual reflection is. It is calculated from the counting statistics, background height and the number of times the reflection was measured. The bigger the sigma, the more imprecise the reflection. Therefore, the bigger the I/σ, the better the data quality. Moreover, I/σ is also used as a criterion for resolution cut-off.

(b) R_{merge}:

R_{merge} is a measure of the difference between symmetry-related reflections that should have theoretically identical intensity.

$$R_{merge} = \Sigma_h | I - \langle I \rangle | / \Sigma_h | I |$$

$\langle I \rangle$ is the mean intensity of all observations of reflection, h=hkl

Traditionally, if the R_{merge} is < 5%, the data are considered as good. If the R_{merge} is between 5%~10%, the data are still useful. But if the R_{merge} is > 10%, the data require further investigation, which may just mean reprocessing. However, R_{merge} is not a very sensitive criterion of the data quality. As the redundancy goes up, the final averaged data quality is improved, but the R_{merge} also goes up. Furthermore, if the space group has high symmetry, the R_{merge} is bigger although the data quality is the same as that of a lower symmetry space group.

Furthermore, R_{merge} can be manipulated by rejecting weak reflections, cutting the higher resolution data off and omitting the partially recorded reflections. In these cases, the quality of data becomes worse although the R_{merge} seems to improve. Therefore, most people prefer to use I/σ as the main criterion to assess the data quality.

2.5.4 Results

2.5.4.1 Data processing parameters

Table 2-E. Data processing parameters for BLG and CYP3

(1) BLG

Ligand	Mosaicity	Spot size (x,y) (mm)	Profile fitting radius (mm)	Crystal orientation (x, y,z)
Native	0.652	(0.45, 0.45)	30	(16.2, -118.4, -98.9)
Palmitate[1]	0.689	(0.5, 0.5)	27	(-19.53, 144.20, -80.32)
Palmitate[2]	0.328	(0.5, 0.5)	27	(0.20, 8.81, 24.75)

palmitate[1] : data collection of BLG co-crystallised with palmitate (first preparation, section 2.3.2.2.)
palmitate[2]: data collection of BLG complexed with palmitate (second preparation, section 2.3.2.3.)

(2) CYP3

Ligand	Mosaicity	Spot size (x,y) (mm)	Profile fitting radius (mm)	Crystal orientation (x, y,z)
Native	0.607	(0.45, 0.45)	27	(70.3, 50.48, -145.0)
1.2mM Ala-Pro	0.757	(0.45, 0.40)	27	(-74.7, 166.3, -169.9)
6mM Ala-Pro	0.614	(0.45, 0.5)	27	(112.7, -168.2, -162.4)
18mM Ala-Pro	0.610	(0.50, 0.45)	27	(-117.3, -28.4, 171.48)
36mM Ala-Pro	0.827	(0.50, 0.45)	27	(-113.8, 33.3, -178.2)
60mM Ala-Pro	1.283	(0.45, 0.45)	27	(149.0, 156.0, 162.1)
120mM Ala-Pro	0.612	(0.45, 0.40)	27	(122.9, 20.7, -23.4)

2.5.4.2 Space group determination

(1) BLGY

(i) Lattice selection:

Lattice	P cubic	I cubic	F cubic	P rhombohedral	P hexagonal	P tetragonal	I tetragonal
distortion index (%)	30.72	31.84	30.83	34.81	3.80	9.05	9.03

Lattice	P orthorhombic	**C orthorhombic**	I orthorhombic	F orthorhombic	P monoclinic	C monoclinic	P triclinic
distortion index (%)	9.05	**0.06**	9.02	8.20	0.05	0.04	0.00

P: Primitive I : I centred F : F centred C: C centred

From the table above, it is shown that the lattice type is C orthorhombic

(ii) Laue group determination:

Laue class	$C222_1$
χ^2	0.996

From the table above, it is shown that the Laue group is $C222_1$

(iii) Systematic absence:

The systematic absence of the data of BLGY is listed in the appendix 2-1. It is shown the reflection condition of BLGY is (0 0 2n).

Therefore, from (i) (ii) and (iii), it is known that the space group is $C222_1$.

(2) BLG co-crystallised with palmitic acid

(I) Lattice selection:

Lattice	P cubic	I cubic	F cubic	P rhombohedral	**P hexagonal**	P tetragonal	I tetragonal
distortion index (%)	52.68	41.02	34.46	8.78	**0.09**	7.81	4.22

Lattice	P orthorhombic	C orthorhombic	I orthorhombic	F orthorhombic	P monoclinic	C monoclinic	P triclinic
distortion index (%)	7.81	0.08	4.22	4.42	0.07	0.05	0.00

P: Primitive I : I centred F : F centred C: C centred

From the table above, it is shown that the lattice type is P hexagonal

(ii) Laue group determination:

Laue class	$P3_1$	$P3_112$	**$P3_121$**	$P6_1$	$P6_122$
χ^2	1.020	21.865	**1.025**	21.890	20.58

From the table above, it is shown that the Laue group is $P3_121$

(iii) Systematic absence:

The systematic absence of the data of BLG co-crystallised with palmitic acid is listed in the appendix 2-1. It is shown the reflection condition is (0 0 3n).

Therefore, from (i) (ii) and (iii), it is known that the space group is $P3_121$ or $P3_221$.

(3) CYP3

(I) Lattice selection:

Lattice	P cubic	I cubic	F cubic	P rhombohedral	P hexagonal	**P tetragonal**	I tetragonal
distortion index (%)	56.63	31.23	33.82	9.79	8.12	**0.10**	6.87

Lattice	P orthorhombic	C orthorhombic	I orthorhombic	F orthorhombic	P monoclinic	C monoclinic	P triclinic
distortion index (%)	0.05	0.07	6.87	7.32	0.04	0.06	0.0

P: Primitive I : I centred F : F centred C: C centred

From the table above, it is shown that the lattice type is P tetragonal

(ii) Laue group determination:

Laue class	$P4_1$	**$P4_12_12$**
χ^2	1.080	**1.003**

From the table above, it is shown that the Laue group is $P4_12_12$

(iii) Systematic absence:

The systematic absence of the data of CYP3 is listed in the appendix 2-1. It is shown the reflection conditions are (0 0 4n) and (2n 0 0).

Therefore, from (i) (ii) and (iii), it is known that the space group is $P4_12_12$ or $P4_32_12$

2.5.4.3 Data processing results

Table 2-F. Data processing results of BLG and CYP3

(1)BLGY

Data collection statistics	BLGY native
Space group, $C222_1$	a = 55.188 Å b= 79.73Å c = 66.684Å
Resolution range	23-2.0Å
Number of reflections measured	131207
Number of unique reflections	10343
Completeness	99.7%
Multiplicity	12.68
R_{merge} (outer shell)	5.1% (20.8%)
I/σ (I) (outer shell)	27.22 (6.30)

(2) BLG complexed with palmitate

Data collection statistics	Palmitate[1]	Palmitate[2]
Space group, $P3_221$	a = b = 53.48 Å c = 111.64Å α = β = 90°, γ = 120°	a = b = 53.52 Å c = 111.37Å α = β = 90°, γ = 120°
Resolution range	20-2.5 Å	25-2.3 Å
Number of reflections measured	97142	48843
Number of unique reflections	6888	8763
Completeness	99.9%	97.8%
Multiplicity	14.1	5.57
R_{merge} (outer shell)	7.3% (36.6%)	6.3% (37.6%)
I/σ (I) (outer shell)	21.2 (3.95)	17.38 (2.57)

palmitate[1] : data collection of BLG co-crystallised with palmitate (first preparation, section 2.3.2.2.)
palmitate[2]: data collection of BLG complexed with palmitat (second preparation, section 2.3.2.3.)

(3) CYP3

Data collection statistics	Native	1.2mM Ala-Pro	6mM Ala-Pro
Space group, $P4_12_12$	a = b = 60.60 Å c = 123.1Å α = β = γ = 90°	a = b = 60.87 Å c = 122.99Å α = β = γ = 90°	a = b = 60.74 Å c = 122.72Å α = β = γ = 90°
Resolution range	20-1.8 Å	25-2.1 Å	25-2.0 Å
Number of reflections measured	145422	165474	189319
Number of unique reflections	22053	14262	16311
Completeness	97.0%	98.9%	99.7%
Multiplicity	6.59	11.60	11.60
R_{merge} (outer shell)	8.3% (27.8%)	9.9% (36.2%)	8.1% (31.5%)
I/σ (I) (outer shell)	11.7 (2.60)	12.50 (3.36)	14.88 (3.62)

Data collection statistics	18mM Ala-Pro	36mM Ala-Pro	60mM Ala-Pro
Space group, $P4_12_12$	a = b = 61.03 Å c = 122.7Å α = β = γ = 90°	a = b = 61.42 Å c = 122.47Å α = β = γ = 90°	a = b = 61.05 Å c = 122.67Å α = β = γ = 90°
Resolution range	20-1.9 Å	20-1.9 Å	25-2.0 Å
Number of reflections measured	146663	224150	270345
Number of unique reflections	19030	19185	16443
Completeness	98.6%	99.5%	95.8%
Multiplicity	7.70	11.60	16.44
R_{merge} (outer shell)	6.7% (34.6%)	7.2% (31.3%)	7.1% (33.2%)
I/σ (I) (outer shell)	13.88 (2.34)	14.89 (2.90)	14.44 (2.09)

Data collection statistics	120mM Ala-Pro
Space group, $P4_12_12$	a = b = 61.03 Åc = 122.66Å α = β = γ = 90°
Resolution range	20-1.9 Å
Number of reflections measured	152448
Number of unique reflections	19230
Completeness	97.6%
Multiplicity	7.92
R_{merge}[1] (outer shell)	7.9% (33.2%)
I/σ (I)[2] (outer shell)	12.85 (2.13)

[1] $R_{merge}=\Sigma_h | I- \langle I \rangle | / \Sigma_h | I |$, where $\langle I \rangle$ is the mean intensity of all observation of reflection h=hkl

[2] σ (I) is the SD of the measured intensity

2.6 Molecular Replacement

2.6.1 Introduction

Molecular replacement (MR) has been proven to be an efficient method to solve protein structures by using knowledge of similar structure. This method is based on the theory that two similar structures in different crystallographic environments, e.g. different space groups, would be expected to have similar diffraction patterns albeit sampled differently. Molecular replacement compares the Patterson function of an unknown structure with the calculated and rotated/translated Patterson function of a search model. Electron density maps calculated with phases with the rotated search model and the unknown structure factors should then reveal the unknown structure. The aim of molecular replacement is to determine the position of the search model within the crystal unit cell. The position of the search model is defined by six parameters, three rotational and three translational parameters. The rotational parameters are specified in terms of Eulerian angle (α, β, γ) while the translational parameters are specified by fractional coordinate (x, y, z). The rotational parameters take values within the range: $0° \leq \alpha \leq 360°$, $0° \leq \beta \leq 180°$, $0° \leq \gamma \leq 360°$ while the translational parameters take values within the range: $0 \leq x \leq 1$, $0 \leq y \leq 1$, $0 \leq z \leq 1$.

The mathematical expression of molecular replacement is as follows:

$r = Rr^0 + T$

where

R is the rotation matrix
T is the translation vector
r, specified by the atomic vectors, is the current position of the search model
r^0 is the initial position of the search model. Conventionally, the initial position is placed where the centre mass of the model is at the origin and the principal axes of the model are parallel to the orthogonal axes. Therefore, by default, $r^0=0$.

The rotation function determines the rotational parameters by comparing the Patterson function of diffraction data with that of the model. The translation function determines the vector necessary to translate the oriented model obtained from the rotation function in order to obtain the best fit with the diffraction data. The standard

procedure of molecular replacement is firstly to calculate the rotation function followed by the translation function.

The success of molecular replacement depends on several factors: completeness and accuracy of the model, quality and completeness of diffraction data, similarity between model and the actual structure and the selected resolution range of the data. In this thesis, AMoRe (Navaza and Saludjian, 1997) was used in molecular replacement calculations and will be the main program discussed. Other molecular replacement packages are available, for example, X-PLOR (Brunger, 1990a) and are reviewed by Tickle and Driessen (1996).

2.6.2 The rotation function

The rotation function is based on the principle that similar molecules will have similar sets of intramolecular vectors. Therefore, the orientation of the model with respect to the actual structure could be determined by overlapping these two intramolecular vector sets.

The rotation function was first developed by Rossmann and Blow (1962) with the mathematical expression:

$$RF(\Omega) = \int_U P_{cry}(r) \times P_{model}[\Omega r] dV \qquad \text{Equation (1)}$$

where P_{cry} is the Patterson function of the actual molecule
P_{model} is the Patterson function of the search model
Ω is the 3 x 3 rotation matrix described by three angles (e.g. Euler angle)
r is the integration variable
U is the region of integration, usually spherical, centred at the origin.

The criterion of fit of the rotation function is defined by the product between these two Patterson functions, P_{cry} and P_{model}. Therefore, the maximum value of RF (Ω) is the most likely solution for the rotation function as the best fit of P_{cry} and P_{model} will give the maximum RF (Ω).

The region of integration should include the maximum number of intramolecular vectors and the minimum number of intermolecular vectors. In order to make the best use of the mathematical properties of the rotation group, the region of integration is usually spherical. The radius of the integration sphere is commonly 75-80% of the model diameter.

The rotation function can be calculated in real space or in reciprocal space. In the real space rotation function (equation (1)), the Patterson function of the search model is presented as a map of an array of intramolecular vectors and is compared with the Patterson function of diffraction data. In the reciprocal space rotation function (equation (2)), the Fourier transform of the model is compared with diffraction data.

$$RF(\Omega) = \sum |F_{cry}(\mathbf{h})|^2 |F_{model}(\Omega^t \mathbf{h})|^2 \qquad \text{Equation 2}$$

where F_{cry} is the structure factors from the diffraction data
F_{model} is the calculated structure factors of the search model
$\mathbf{h}$ =hkl is the indices of the diffraction pattern
Ω^t is the reverse of the rotation matrix Ω

These two rotation functions are, in principle, equivalent. The reciprocal space rotation function could be converted from the real space rotation function by Fourier transform and application of Parseval's theorem.

2.6.3 The translation function

The translation function was first introduced by Crowther and Blow (1967) with a similar mathematical expression to that of rotation function:

$$TF(v) = \int_V P_{cry}(u)\, P_{model}(u,t)\, du \qquad \text{Equation 3}$$

where P_{cry} is the Patterson function of the actual molecule
P_{model} is the Patterson function of the search model
v is the intermolecular vector between the two symmetry-related molecules
u is the integration variable
V is the region of integration
t is the unknown translation vector of the search model

Again, the maximum value of TF(v) is the most likely solution for the translation function as the best fit of P_{cry} and P_{model} will give the maximum TF(v).
Once the rotation parameters have been obtained from the rotation function, the oriented search model has to be translated to the optimal position in the crystal unit cell by calculating the translation function.
Similar to the rotation function, the translation function is to determine the translational parameters by comparing the Patterson function of diffraction data with that of the model. However, intermolecular vector sets are used in the calculation of

the translation function whereas the intramolecular vector sets are used in the calculation of the rotation function.

2.6.4 Molecular replacement using AMoRe

AMoRe (Navaza and Saludjian, 1997) is an automated molecular replacement program implemented in CCP4. It consists of five steps: format of the input data for use in later steps, fast calculation of structure factors by continuous Fourier transform of the model electron density, calculation of the rotation function, calculation of the translation function and rigid-body refinement of the orientated model obtained from molecular replacement. These five steps were performed in order by the following programs:

(a) Sorting:

The program SORTING is to sort and pack the input data (H, K, L, Fobs) to an AMoRe readable format. Usually, data with a low resolution cut-off to 10Å are used as data lower than 10Å mainly contains information about solvent.

(b) Tabling:

The program TABLING is used to translate the model so that the centre of mass is at the origin and performs fast computation of structure factors of the search model. In AMoRe, a very efficient algorithm is applied to calculate the structure factors of the model in terms of the continuous Fourier transform. The output from running TABLING produces the Fourier coefficient of the search model which is used as the input into the programs ROTING and TRAING which will be discussed later.

(c) Roting:

The program ROTING is used to calculate the rotation function. A modified fast rotation function is used in AMoRe. The fast rotation function was introduced by Crowther (1972) and is an expansion of reciprocal space Patterson function in terms of spherical harmonics. By this method, the entire rotation function could be calculated by a single Fourier transformation and consequently the calculation of rotation function is faster.

The output of this program is the rotational solution given in terms of the Eulerian angles (α, β, γ).

(d) Traing:

The program TRAING is used to calculate the translation function. The translation function devised by Crowther and Blow is applied in AMoRe. The input is the rotation solution specified by Eulerian angles. After calculation of translation function, the translation solutions in terms of fraction coordinate (x, y, z) along with correlation coefficient and initial R-factor are given. Good solutions will have high correlation coefficients and low R-factors. The correlation coefficient is a measure of the difference between the structure factors calculated from the model and those obtained from the experiment. Therefore, the correlation coefficient is a measure of the agreement of the model and actual structure. The higher the correlation coefficient, the better agreement between the model and the actual structure.

(e) Rigid-body refinement:

After translation function, the oriented model is subject to rigid-body refinement in order to obtain the maximum correlation coefficient and consequently, the optimum position.

2.7 Refinement and Model building

2.7.1 Introduction

As the structures solved by MIR (multiple isomorphous replacement) and MR (molecular replacement) are usually inaccurate and contain a large number of errors, refinement of the structure is necessary to bring observed and calculated intensities or structure factor amplitudes as close as possible. Refinement is the process of adjusting the model to improve the agreement between the observed and calculated structure factors. For successful least squares refinements, the number of the observed reflections should be significantly greater than the number of parameters. However, in protein structures, the ratio of observed reflections to refinable parameters (x,y,z,B) is usually very low because of the loss of high resolution data due to disorder in the crystal and high solvent content in the crystal. In order to overcome this problem, constraints or more usually nowadays restraints are applied in the refinement procedure (Glusker et al., 1994a). Constraints reduce the number of

parameters by expressing several parameters in terms of a single parameter. For example, a single set of parameters (x,y,z, θ_1, θ_2, θ_3) is used to represent the position and orientation of the whole group of atoms. Restraints can be considered as additional observed information usually derived from the stereochemical information such as bond length, bond angle, torsion angle, bond planarity and non-bonded information such as van der Waals interactions. Thus, the restraint refinement will not only increase the number of observations in calculations but also help ensure that the stereochemistry of the molecule is correct. Therefore, the overall function being minimised contains two terms: the crystallographic term defined as the square of the difference between the observed structure factor amplitude ($|F_o|$) and calculated structure factor amplitude ($|F_c|$) and the stereochemical term defined as the square difference between the actual value and ideal value of the geometric parameters, like bond length and bond angle (McRee, 1993). These ideal values are obtained from the highly refined small molecule structures.

$$E_{total} = \omega_a \Sigma E_{crystallographic} + \omega_b \Sigma E_{stereochemical}$$

$$E_{crystllographic} = [\ |F_o(hkl)| - F_c(hkl)|\ ]^2$$

$$E_{stereochemical} = \Sigma_{bond\ lentgh}(\ A-A_o)^2 + \Sigma_{bond\ angle}\ (\ B-B_o)^2 + \Sigma_{\ torsion\ angle}(\ C-C_o)^2 + \Sigma_{non\text{-}bonded\ contacts}\ (D-D_o)^2 + \Sigma_{planar\ groups}\ (\ E-E_o)^2 + \Sigma_{\ chiral\ volumes}(\ F-F_o)^2$$

where
A, B, C, D, E, F : the actual value
A_o B_o C_o D_o E_o F_o : the ideal value

ω_a and ω_b are the weights of the crystallographic term and stereochemical term, respectively. A good refinement program should be able to apply the appropriate weighting scheme in the refinement procedure as the weight can affect the result of the refinement dramatically. If the weight of the crystallographic term is too high, the stereochemical geometry of the structure will be distorted. If the weight of the stereochemical term is too high, it will slow down the convergence of the refinement and result in a poor fit to the observed x-ray data.

A least squares refinement program aims to minimise the overall function by calculating the shift in coordinate or temperature factor that will give the lowest value.

Because the least squares process calculates a shift from an input position, if that position is far from the true one, no satisfactory shift can be determined and no improvement result. In other words, least squares refinement requires a reasonably accurate input model. To overcome this short-coming, following each round of refinement, an electron density map is generated and displayed in the computer graphic system to ensure the model is well-fitted in the density map. Manual model building, including moving atoms, changing conformation, insertion of amino acids and addition of ligands may be required if the fit is poor. Two types of electron density map are usually used for model building, (2Fo-1Fc) and (Fo-Fc) maps. (2Fo-1Fc) (or (3Fo-2Fc)) map is like as the sum of the electron density of a model and a difference electron density map. It is usually used to check if each part of the model fits well into the map (Glusker *et al.*, 1994b). The other one is the (Fo-Fc) map, the so-called difference map. It is used to highlight the errors in a trial structure, to find out parts of the structure that have not yet been identified or located and to recognise the movement of residues. Once the native structure has been solved, the difference map is also a very important tool in studies of complexed structures since it can be used to identify the location of a ligand or substrate soaked into a crystal or co-crystallised with the protein.

Several algorithms have been established to refine the protein structures. In general, they can be classified into conjugate gradient minimisation (Jack and Levitt, 1978), constrained or restrained least-square refinement (Hendrickson, 1985) and simulated annealing (Brunger *et al.*, 1987). Conjugate gradient minimisation and simulated annealing were used in X-PLOR to refine the macromolecule structures. Least-square refinement and conjugate gradient minimisation were applied in SHELX97 to refine the small molecule and macromolecule structures, respectively. These two programs are used in the work of this thesis and will be discussed in the following sections.

2.7.2 Refinement program—X-PLOR

2.7.2.1 Introduction

X-PLOR (Brunger *et al.*, 1992) was initially used for the refinement of the native BLGY structure and structure of palmitate bound to BLG.

This program contains a variety of refinement protocols, including rigid-body refinement, positional refinement, B-factor refinement and simulated annealing. An energy minimisation procedure was used in X-PLOR where the crystallographic and stereochemical terms are treated as functions of energy.

(a)Rigid-body refinement:

In rigid-body refinement, the structure of protein is broken into several rigid bodies. These rigid bodies can be domains, polypeptide chains or monomers. Three rotational and three translational parameters are assigned to each rigid body and these parameters are then refined to minimise the total energy function. The molecular conformation of each rigid body remains the same when the six rotational and translational parameters for each are refined. In rigid-body refinement, the ratio of observed reflections to refinable parameters is high as only six parameters have to be refined for each rigid body.

The data of lower resolution, i.e. lower than 3Å, was used in order to increase the radius of convergence. The radius of convergence is d/4 in theory, where d is the highest resolution data included in the refinement procedure (Jack and Levitt, 1978).

(b)Positional refinement:

In X-PLOR, the conjugate gradient minimisation method (Jack and Levitt, 1978) is used in conventional positional refinement. The overall energy function is minimised by shift in coordinates to get the lowest energy. As mentioned in section 2.7.1, it is essential to use the appropriate weight to balance the contribution from the crystallographic term and stereochemical term. In XPLOR, the program CHECK can determine the appropriate weight by running a molecular dynamics calculation.

(C)*B-factor refinement*:

B-factor refinement is performed independently from positional refinement in X-PLOR. There are three types of B-factor refinements in X-PLOR: an overall B-factor,

a grouped B-factor and individual B-factor refinement. A single B-factor is calculated for the entire molecule in overall B-factor refinement and this single B-factor is an indicator of the mobility of the whole molecule. In grouped B-factor refinement, single B-factor is applied to the atoms of the same group while, in individual B-factor refinement, individual B-factor is calculated for each atom. The choose of the type of B-factor refinement used depends on the resolution of the data. Usually, overall B-factor refinement or grouped B-factor refinement is used in the early stage of refinement whereas individual B-factor refinement is used when higher resolution data is included in the latter stage of refinement.

(d)Simulated annealing:

Simulated annealing has a wider radius of convergence than least-square refinement and conjugate gradient minimisation. Therefore, simulated annealing can reach global minima whereas least-square refinement or conjugate gradient minimisation is easily trapped in a local minimum. In simulated annealing, the structure is heated to high temperature, usually 2000K to 3000K, and then annealed the structure by slowly reducing the temperature. The "temperature" here is a control parameter that determined whether the structure can escape the local minima instead of a physical temperature. Therefore, very high temperature maybe has to be used if the barriers between local minimum are rather large.

2.7.2.2 Refinement procedure

The input PDB file was subjected to a program call GENERATE (also in X-PLOR) and two output files were then obtained. One is the coordinate file with a header in the beginning of the file. The other is the structure file (.psf file) containing structural and stereochemical information, for example, bond length, bond angle, etc.

5% of all reflections randomly selected by the program SETUP FREE R to monitor the refinement procedure in order to avoid over-refinement are flagged as R_{free} set and excluded in all refinement.

Prior to any refinement, the program CHECK should be used to calculate the appropriate weight. This program also calculates an initial R-factor before refinement was carried out.

Rigid-body refinement was performed firstly and was followed by simulated annealing. Before simulated annealing using the program SLOWCOOL, the program PRESTAGE was used to minimise the energy of structure. Following simulated annealing, positional refinement and then B-factor refinement were performed. After each round of refinement, electron density map was calculated and displayed in O (Jones *et al.*, 1991). The map generated in X-PLOR has to be converted into O format by a program called MAPPAGE (Jones, 1992) from the RAVE suite.

2.7.3 Refinement program—SHELX97

2.7.3.1 Introduction

SHLEX97 (Sheldrick and Schneider, 1997) is a multi-program used to carry out refinement of all of the structures discussed in this thesis. This program uses a conventional structure factor calculation rather than a FFT (fast Fourier transform) summation usually applied in other programs, like X-PLOR. The FFT method involves small approximations whereas the conventional structure factor calculation method is more accurate although it takes more CPU running time. The other difference in SHELX97 is this program always refines against the square of the structure factor amplitude ($|F|^2$) instead of the structure factor amplitude ($|F|$) since from experience with small molecules, the refinement against $|F|^2$ is better than $|F|$. SHLEX97-SHELXL is an easy-to-use program, special for macromolecules, even at moderate resolution around 2.5Å and only two input files are required. One is the reflection data file (.hkl) containing h,k,l,F^2 and σ (F^2) while the other is the instruction file (.ins) including the appropriate stereochemical restraints. The reflection file from DENZO and the instruction file from PDB format file are generated in SHELXPRO, an interactive program interfacing with the other protein structure programs.

2.7.3.2 Refinement procedure

In SHELXL, positional and B-factor refinement are performed in the same cycle. Usually, the data between 10Å to the highest resolution are used for refinement. 5% of all reflections randomly selected by SHELXPRO to monitor the refinement

procedure in order to avoid over-refinement are flagged as the R_{free} set and excluded in all refinement and map calculation. Following each round of refinement, (2Fo-1Fc) or (Fo-Fc) map is generated in SHELXPRO and displayed in the model building program O. H_2O molecules are added by SHELXWAT, an automatic H_2O adding program implemented in SHELX97. After checking H_2O generated by SHELXWAT, H_2O molecules with high B-factor and unfavourable geometry relationship with protein are deleted.

3. Structure determination of BLGY

This chapter will describe the crystallographic work on the BLGY structure. Section 3.1 is to determine and refine the structure. Section 3.2 is to examine the detail of the structure. Section 3.3 and 3.4 is to compare the BLGY structure with the BLGZ and BLGX structures, respectively. Section 3.5 is to discuss the problems encountered in the BLGY structure.

3.1 Molecular replacement

The first lattice Y structure was solved to 6.0Å resolution by Green and co-workers in 1979 (Green *et al.*, 1979). Later, in 1986, a higher resolution structure to 2.8Å resolution was published by Papiz *et al* (1986). However, successful refinement of this structure could not be achieved. Therefore, in 1990, Hambling collected much higher resolution data to 1.8Å (Hambling, 1990) and tried to refine the structure solved by Papiz. This data, phased to 2.8Å resolution, was used to calculate an electron density map and this map appeared better than that obtained by Papiz although the density for residues 55 to 69 and 150 to 162 was still ill-defined. A model was rebuilt into the 2.8 Å (2Fo-Fc) map and this rebuilt model was then subject to further refinement. However, further refinement of this model failed as the R-factor fluctuated around the mid 40's to 2.8Å resolution and this refined model had unusual geometry with many residues in the disallowed regions of the Ramachandran plot. One year later, manual-building of the Hambling model, with geometry restraints being applied regularly, was carried out by McAlpine (1991). The R-factor reduced to 34% to 2.8Å resolution and acceptable geometry was obtained. However, further refinement failed to reduce the R-factor below 34% although many approaches were tried. After McAlpine, Dr. Igor Polikarpov in Edinburgh collected a new data set in collaboration with Dr. Sean McSweeney in Daresbury and tried to re-solve the lattice Y structure by molecular replacement using the lattice X structure, which was later published by Dr. Brownlow in 1997 as mentioned in section 1.1.4.2.

However, this structure also revealed some problems as the two terminal regions and several loops had extremely high B-factors and questionable geometry (Polikarpov, 1995)

After the project of this thesis started, independently, Dr. Maria Bewley in professor Ted Baker's laboratory in New Zealand re-collected the data of lattice Y form at 1.8Å and re-solved the structure by multiple isomorphous replacement.

The newly solved structure, requiring further refinement was kindly provided by Dr. Maria Bewley and used as the initial model for the further refinement described in this thesis. This original model had some missing residues in the N-terminus and the GH loop: residue 1 to 2 in the N-terminus and residue 111 to 114 in the GH loop.

3.1.1 Molecular replacement using the lattice Y model provided by Dr. Maria Bewley

The structure solved by Dr. Maria Bewley was from data collected at room temperature. However, data used for the work in this thesis were collected at low temperature (100K). This resulted in a decrease of the unit cell dimensions from a=55.5Å, b=82.1Å, c=66.8Å to a=55.19Å, b=79.72Å, c=66.68Å, a contraction largely along the b axis, which is also the molecular dyad. Thus, molecular replacement and rigid-body refinement were performed firstly to ensure the position of the model in the unit cell.

The model (space group $C222_1$; a=55.5Å, b=82.1Å, c=66.8Å, $\alpha=\beta=\gamma=90°$) was subjected to molecular replacement using the program AMORE. It is noted that the space group here is the standard setting $C222_1$ instead of permuted cell, $B22_12$, used by Green *et al* (1979) and Papiz *et al* (1986). The input values for variables used with AMORE are listed in Table 3-A.

Table 3-A. Input values for variables used with the AMORE program

Resolution (Å)	10-4
P1 box size (Å,°)	(75, 75, 75, 90, 90, 90)
Radius of integration (Å)	18
Angular step (°)	1.0

The top six peaks of the rotation function are listed in Table 3-B.

Table 3-B. The top six peaks of the rotation function using the model from Dr. Maria Bewley as the search model

Peak	(α, β, γ)	Correlation coefficient (%)	Sigma (σ)
1	(359.00, 0.00, 0.00)	36.2	7.24
2	(179.00, 0.00, 0.00)	36.2	7.24
3	(0.00, 180.00, 179.00)	36.2	7.24
4	(0.00, 180.00, -1.00)	36.2	7.24
5	(73.60, 169.22, 101.54)	19.4	3.88
6	(106.40, 10.78, 281.55)	19.4	3.88

After checking with the CCP4 program ROTMAT (Dodson,1997), which generates the symmetry equivalents, it became apparent that peaks 1,2, 3 and 4 are symmetry equivalent.

The maximum peak (7.24σ) in the rotation function and next highest peak (3.88σ) were used to calculate the translation function. After the translation search, the maximum peak (α=359.00, β=0.00, γ=0.00) gave a distinct solution with a high correlation coefficient of 64.3% and a low R-factor of 38.8%. The next solution has a lower correlation coefficient of 45.3% and a higher R-factor of 46.5%.

The top five solutions of translation function using this maximum peak are listed in Table 3-C.

Table 3-C. The top five solutions of translation function using the maximum peak obtained from rotation function (α=359.00, β=0.00, γ=0.00)

Solutions of translation function	Fractional coordinates (Tx, Ty, Tz)	Correlation coefficient (%)	R-factor (%)
1	(0.2429, 0.1517, 0.3556)	64.3	38.8
2	(0.2433, 0.3976, 0.3556)	45.3	46.5
3	(0.2400, 0.2690, 0.3528)	42.5	47.7
4	(0.4854, 0.4036, 0.3517)	41.6	49.1
5	(0.2452, 0.0190, 0.3566)	41.5	47.3

The packing of the molecules in the unit cell was checked by the program PACMAN (Kleywegt *et al.*, 1996) and also displayed in O (Jones *et al.*, 1991). Both programs showed that the packing of the first solution was reasonable with no symmetry-related molecules clashing into each other.

It was obvious that the first solution was correct since it had a significantly higher sigma value, higher correlation coefficient and lower R-factor when compared to the second solution. The packing of this solution was also reasonable.

As well as peak 1 of table 3-B, peak 5 (α=73.60, β=169.22, γ=101.54) obtained from the rotation function was used to calculate the translation function. However, unlike peak1, it failed to give a distinct solution. In addition, the solutions that it did give had much lower correlation coefficients (approximately 21%), and much higher R-factors (approximately 55%).

Following the translation function, rigid-body refinement of the optimum molecular replacement solution was performed. This increased the correlation coefficient significantly from 64.3% to 72.1% and reduced the R-factor from 38.8% to 32.9%

The Eulerian angles and fractional coordinates shift to α=359.57, β=-0.24, γ=1.01, Tx=0.2422, Ty=0.1555, Tz=0.3535.

3.1.2 Molecular replacement using a monomer of lattice X

A monomer of the crystal structure of dimeric BLG refined at 1.8Å resolution in triclinic Lattice X (space group P1; a=37.8Å, b=49.6Å, c=56.6 Å, α=123.4°, β=97.3°, γ=103.7°; PDB ID 1BEB) was also used as the search model (Brownlow *et al.*, 1997) in the structure determination of the lattice Y crystal form.

Following the same procedure as in section 3.1.1 and using the same input variables as were applied in section 3.1.1, the rotation function gives the maximum peak (α=33.93, β=119.03, γ=333.81) of 6.04σ and the next highest peak (α=42.60, β=16.08, γ=342.53) of 3.92σ. Both were used to calculate the translation function.

After the translation search, the maximum peak (α=33.93, β=119.03, γ=333.81) gave the best solution with a correlation coefficient of 49.0% and an R-factor of 44.8%

while the next solution had a correlation coefficient of 36.6% and an R-factor of 49.5%. The top five solutions of translation function are listed in Table 3-D.

Table 3-D. The top five solutions of translation function using the maximum peak obtained from rotation function (α=33.93, β=119.03, γ=333.81)

Solutions of translation function	Fractional coordinates (Tx, Ty, Tz)	Correlation coefficient (%)	R-factor (%)
1	(0.2411, 0.1520, 0.3584)	49.0	44.8
2	(0.2413, 0.3941, 0.3556)	36.6	49.5
3	(0.2523, 0.0112, 0.3607)	33.8	50.2
4	(0.2483, 0.3230, 0.3594)	33.2	51.1
5	(0.4784, 0.1476, 0.3573)	31.7	50.2

The packing of the molecules in the unit cell was checked by the program PACMAN (Kleywegt *et al.*, 1996) and also displayed in O (Jones *et al.*, 1991). Both programs showed that the packing of the first solution was reasonable with no symmetry-related molecules clashing into each other.

The first solution was taken as correct since it had a higher sigma value, higher correlation coefficient and lower R-factor when compared to the second solution. The packing of this solution was also reasonable.

The next highest peak (α=42.60, β=16.08, γ=342.53) of the rotation function gave no significant solution and all other solutions had a lower correlation coefficient around 18.0% and a higher R-factor around 55.0%.

Although correct, the best solution obtained using the lattice X structure as the search model was poor compared to that obtained using the lattice Y model provided by Dr. Maria Bewley as the solution from lattice X structure has much lower correlation coefficient and much higher R-factor. Therefore, the molecular replacement solution using the lattice X structure was discarded and further refinement work in this thesis focuses on the lattice Y model provided by Dr. Maria Bewley.

3.1.3 Molecular replacement using the lattice Z model

The recently solved structure of BLG refined at 2.24Å resolution in lattice Z form (space group $p3_221$; a=b=53.96Å, c=112.41 Å, $\alpha=\beta=90.0°$, $\gamma=120.0°$; PDB ID 1BSY) was also used as the search model (Qin *et al.*, 1998a) in the structure determination of the lattice Y crystal form to examine whether better molecular replacement solutions could be obtained from this model.

Following the same procedure as in section 3.1.1 and using the same input variables as were applied in section 3.1.1, the rotation function gives the maximum peak ($\alpha=172.51$, $\beta=48.63$, $\gamma=153.15$) of 5.72σ and the next highest peak ($\alpha=154.07$, $\beta=10.71$, $\gamma=214.44$) of 3.67σ. Both were used to calculate the translation function. After the translation search, the maximum peak ($\alpha=172.51$, $\beta=48.63$, $\gamma=153.15$) gave the best solution with a correlation coefficient of 47.1% and an R-factor of 46.0% while the next solution had a correlation coefficient of 29.8% and an R-factor of 53.3%. The top five solutions of translation function are listed in Table 3-E.

The packing of the molecules in the unit cell was checked by the program PACMAN (Kleywegt *et al.*, 1996) and also displayed in O (Jones *et al.*, 1991). Both programs showed that the packing of the first solution was reasonable with no symmetry-related molecules clashing into each other.

Table 3-E. The top five solutions of translation function using the maximum peak obtained from rotation function ($\alpha=172.51$, $\beta=48.63$, $\gamma=153.15$)

Solutions of translation function	Fractional coordinates (Tx, Ty, Tz)	Correlation coefficient (%)	R-factor (%)
1	(0.2571, 0.1537, 0.1424)	47.1	46.0
2	(0.0801, 0.1594, 0.1444)	29.8	53.3
3	(0.2540, 0.1518, 0.0306)	29.5	51.9
4	(0.2486, 0.1515, 0.4861)	28.4	53.1
5	(0.3502, 0.0328, 0.4727)	28.2	52.8

The first solution was correct since it had a higher sigma value, higher correlation coefficient and lower R-factor when compared to the second solution. The packing of this solution was also reasonable.

The next highest peak (α=154.07, β=10.71, γ=214.44) of the rotation function gave no significant solution and all other solutions had a lower correlation coefficient around 18.4% and a higher R-factor around 55.7%.

The best solution obtained using the lattice Z structure as the search model was poor compared to that obtained using the lattice Y model provided by Dr. Maria Bewley although this solution was also correct. Therefore, the molecular replacement solution using the lattice Z structure was discarded and further refinement work in this thesis focuses on the lattice Y model provided by Dr. Maria Bewley.

3.1.4 Refinement and model building

Following rigid-body refinement, the model was subjected to simulated annealing using data between 10Å to 2.0Å resolution. The slow cooling protocol for refinement by simulated annealing was used (Brunger *et al.*, 1990b). Firstly, the model was heated to 2500K, followed by 50 steps of molecular dynamics integration performed with a time step of 0.5fs. After each molecular dynamics calculation, the temperature was decreased by steps of 25K to the final temperature of 250K. The R-factor decreased from 34% to 32.1% (R_{free} from 42.1% to 40.4%) after simulated annealing. The initial electron density map, (3Fo-2Fc) map, calculated in X-PLOR (Brunger, 1992) and converted in MAPPAGE (Jones, 1992) was displayed in O. The main chains of the core region of BLG were covered by the density map. The core region contained a β-barrel formed by 8 antiparallel β-strands and a α-helix outside the β-barrel. However, the density of the main chains of N-terminus, C-terminus and some disordered loops, such as the AB, CD and GH loops, was rather poor. There was no continuous density to allow the tracing of the main chains of the CD loop and C-terminus. Following simulated annealing, positional refinement and individual B-factor refinement were performed using X-PLOR. This resulted in a slightly decrease of the R-factor from 32.1% to 31.0% (R_{free} from 40.4% to 39.3%). However, there were no improvements in the density of these disordered loops. Therefore, the decision was taken to re-build manually these disordered loops using the program O. The AB loop was re-built firstly since the density of the AB loop was clearer than others. (3Fo-2Fc) and (Fo-Fc) maps were both used to re-build this loop. The (Fo-Fc)

map was used to monitor the movement of the main chain while the (3Fo-2Fc) map was used to identify the orientation of the side chain. The (3Fo-2Fc) map was shown in Figure 3-1.

Manual rebuilding of the AB loop involved the movement of the main chain from residue 32 to 36, the flipping of some peptide bonds and the alteration of some side chain rotamers. Following rebuilding, the model was subjected to positional refinement. After 200 cycles of refinement, the R-factor decreased from 31.0% to 29.7% (R_{free} from 39.3% to 36.1%). The density map was also improved and most atoms of AB loop were fitted into density although the density of some side chains was still very weak. Following refinement, an attempt was made to re-build manually the GH loop. This loop was omitted in the original model. However, in the (3Fo-2Fc) and (Fo-Fc) maps, this loop was identifiable although the density was rather poor. Careful inspection allowed tracing of the main chain through continuous density. After rebuilding, positional and individual B-factor refinement was again performed. However, there was no improvement in the density map. In particular, the density map of the side chains was ill-defined. The geometry of this loop was also poor as the main chains of Pro113 and Ser116 and the side chain of Glu114 and Gln115 adopted unfavourable conformations. Furthermore, there was no significant decrease of the R-factor.

The rebuilt model was subject to further refinement with SHELX97-SHELXL (Sheldrick and Schneider, 1997). After 5 cycles of positional and B-factor refinement, H_2O molecules were added using SHELXWAT. The positions of H_2O were checked manually in O firstly to ensure the location was appropriate and secondly that the H_2O made a reasonable hydrogen bond contact with the protein. A (2Fo-1Fc) map was calculated using SHELXPRO and displayed with O. The refinement and addition of H_2O resulted in a dramatic decrease in the R-factor from 29.7% to 25.9% (R_{free} from 36.1% to 31.5%) and also greatly improved the density map. The density of main chains and side chains of the core region was now well-defined. The density maps of both AB and GH loops were clearer, especially that of the AB loop. The (2Fo-1Fc) map of the AB loop is shown in Figure 3-2.

However, the density maps around the CD loop and the C-terminal regions were still ill-defined.

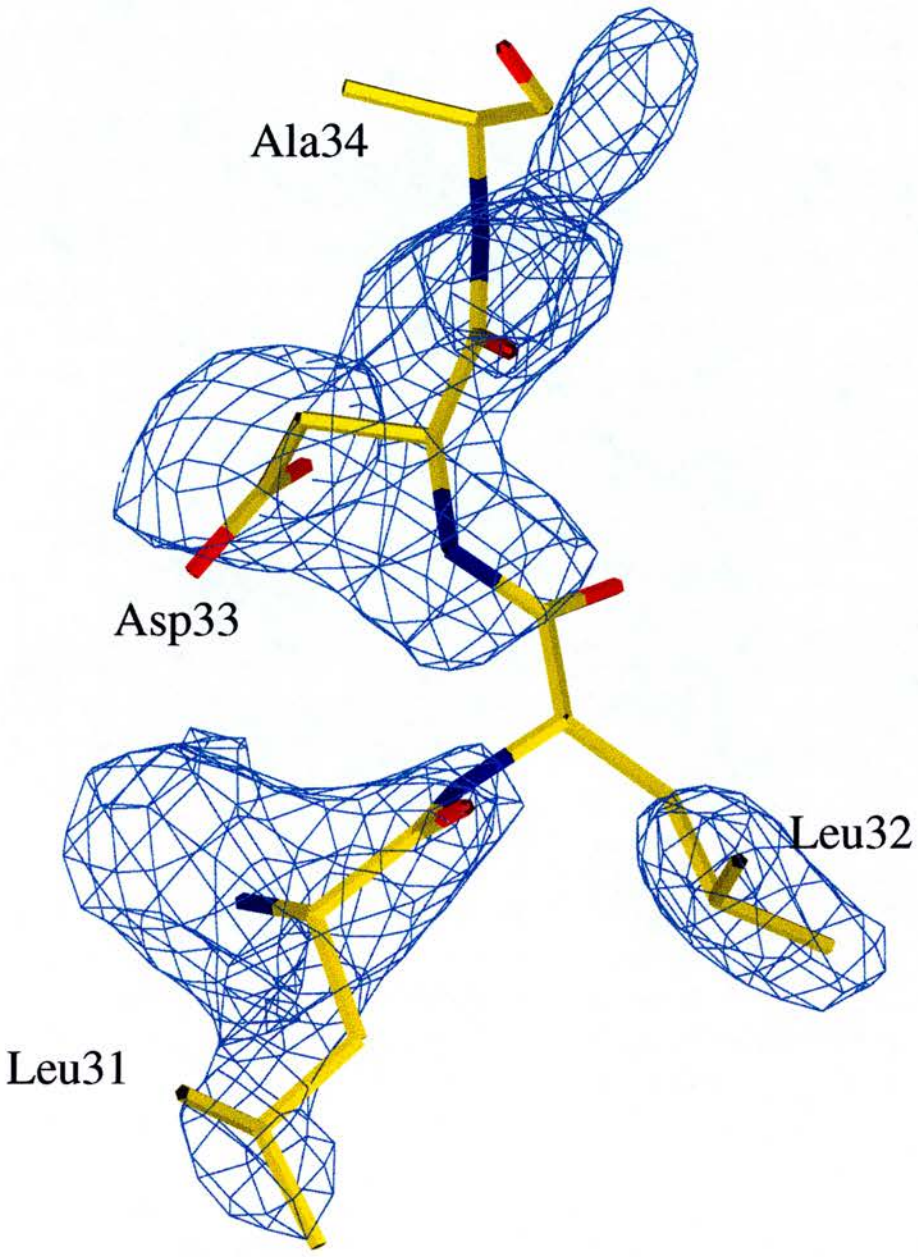


Figure 3-1.The (3Fo-2Fc) electron density map, contoured at 1.2 sigma, of the AB loop. The map was calculated by XPLOR (Brunger, 1992) after simulated annealing refinement and displayed in O (Jones et al., 1991).

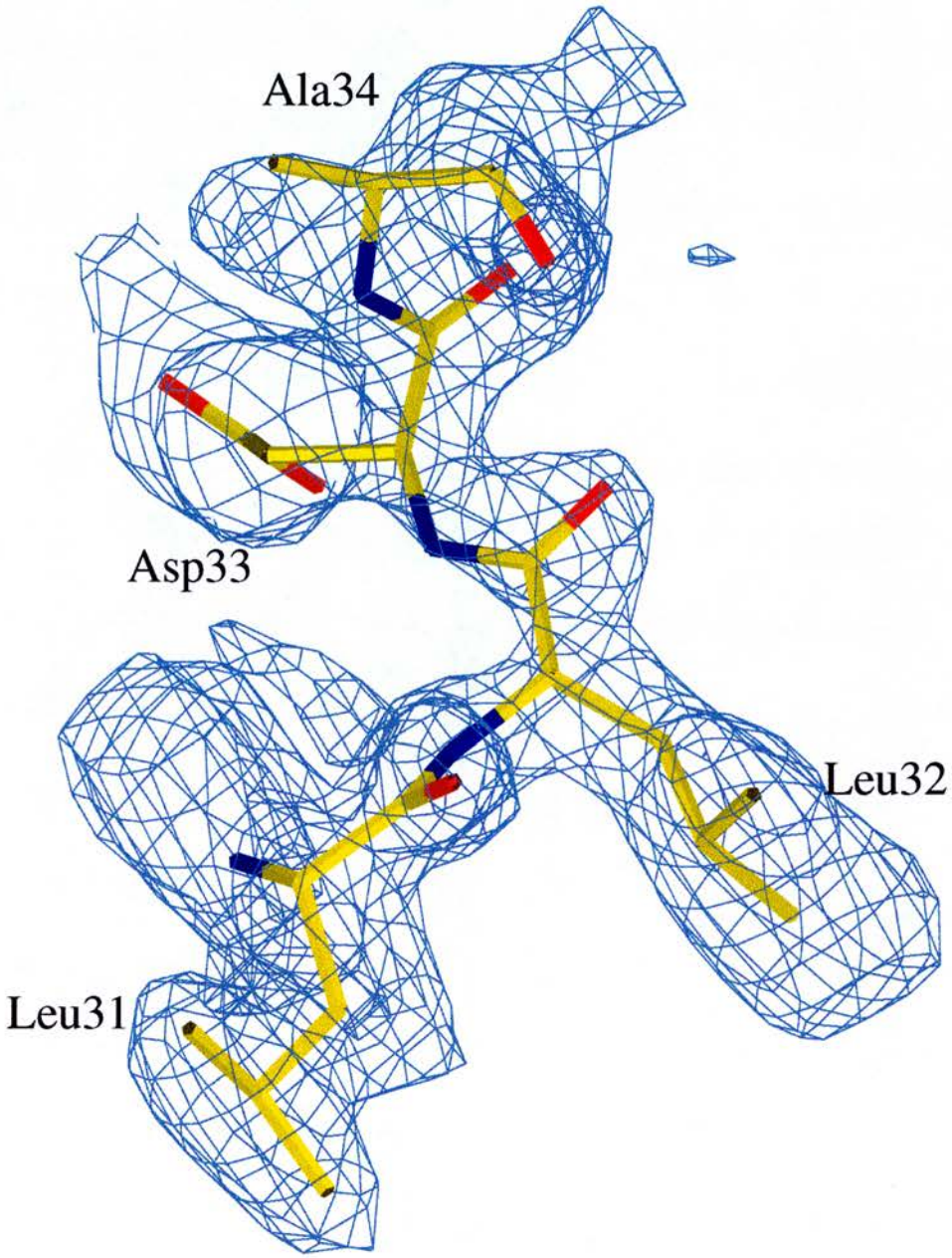


Figure 3-2.The (2Fo-1Fc) electron density map, contoured at 1.2 sigma, of the AB loop. The map was calculated by SHELXPRO (Sheldrick and Schneider,1997) after refinement in SHELX and displayed in O (Jones et al., 1991). This density map shows the great improvement in density of AB loop after the refinement in SHELX compared to Figure 3-1.

The improved density map allowed the GH loop to be built properly. After rebuilding and positional refinement, the main chains of the GH loop were fitted into density. The geometry was better although the density maps of some side chains, such as Glu112, Glu114, and Gln115, were very weak. However, there was no significant decrease in the R-factor and the B-factors of atoms in this region were rather high, the average B-factor being 85Å^2.

It was very difficult to build the CD loop and the C-terminus (see Figure 3-3) since the density was still rather poor.

Several methods were tried to overcome this problem. Firstly, removal of the poorly defined residues was tried. The residues 60 to 66 of the CD loop and residues 152 to 160 of the C-terminal region were removed from the original model and this partial model was then subjected to positional and B-factor refinement. After 15 cycles of positional and B-factor refinement, the R-factor had increased from 25.9% to 27.1% as expected.

Moreover, both (2Fo-Fc) and (Fo-Fc) maps showed no obvious improvement around these two regions. The removed residues were therefore replaced back since removing them seemed to have no beneficial effect on the overall structure.

Secondly, residues from 60 to 66 and from 152 to 160 were converted to be alanine in an attempt to reduce the effect of the side chain atoms in these two regions. After 15 cycles of positional and B-factor refinement, a (2Fo-Fc) map was again generated.

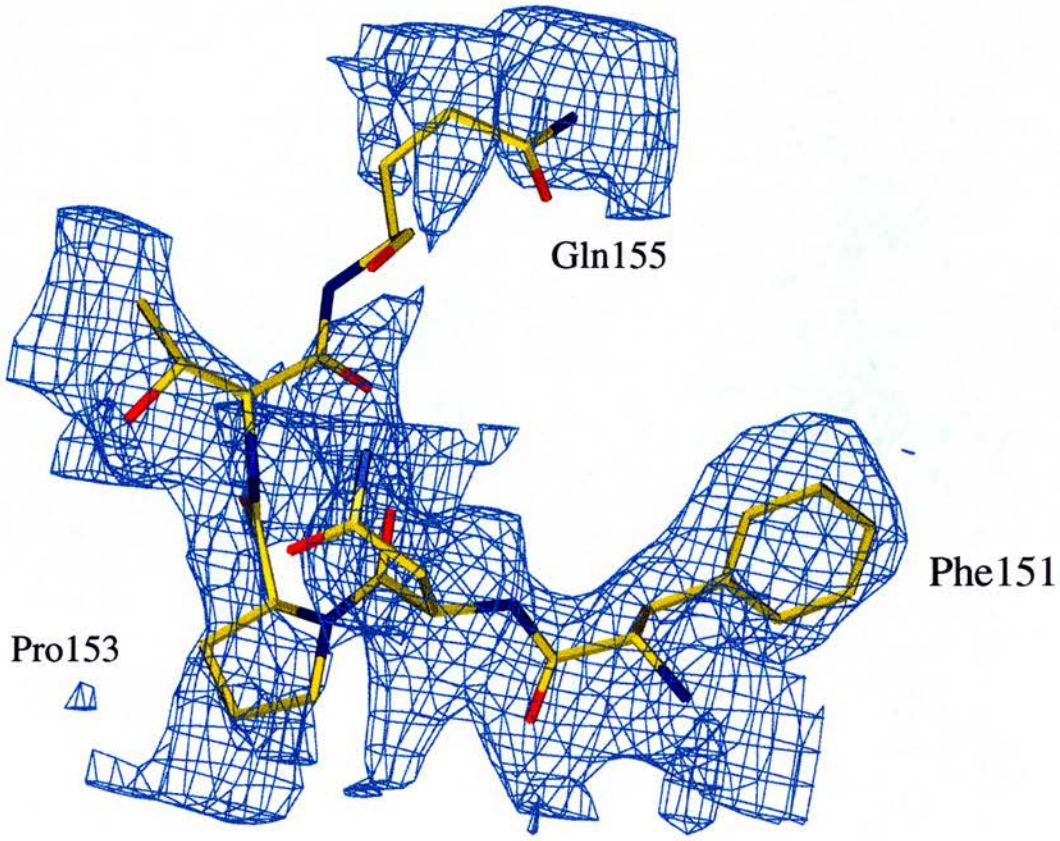


Figure 3-3.The (2Fo-1Fc) electron density map, contoured at 1.0 sigma, of the part of the C-terminus. The map was calculated by SHELXPRO (Sheldrick and Schneider,1997) after refinement in SHELX and displayed in O (Jones et al., 1991).

However, there was still no improvement in the density map. Since there was little effect seen from converting these residues to alanine, the side chains of these residues were replaced into their original positions.

The other approach was to build the C-terminus of the lattice Y structure by using the C-terminal region from residues 152 to 160 of the lattice Z structure (the structure of BLG complexed with palmitate, which will be discussed in chapter 4) as the C-terminus in the lattice Z form was well-fitted into the density (see Figure 3-4) However, the trial also failed since the main chains were out of the density. This result indicated that the C-terminus of the lattice Y structure adopts an alternative conformation to that seen in the lattice Z structure.

Further rebuilding and refinement failed to produce an improved density map and the R-factor continued to fluctuate around 26%.

The current R-factor is 25.7% and R_{free} is 31.0% for all reflections to 2.0Å resolution. The result of the final round of refinement is summarised in Table 3-F.

Table 3-F. Summary of the refinement statistics of BLGY structure.

Summary of the refinement statistics	**BLGY structure**
Resolution range (Å)	10-2.0
R_{final} (F>4σ)	25.7% (25.1%)
R_{free} (F>4σ)	31.0% (30.2%)
rms deviations from ideal geometry	
Bond lengths (Å)	0.005
Angle distance (Å)	0.018
mean B-factor (Å^2)	50.0

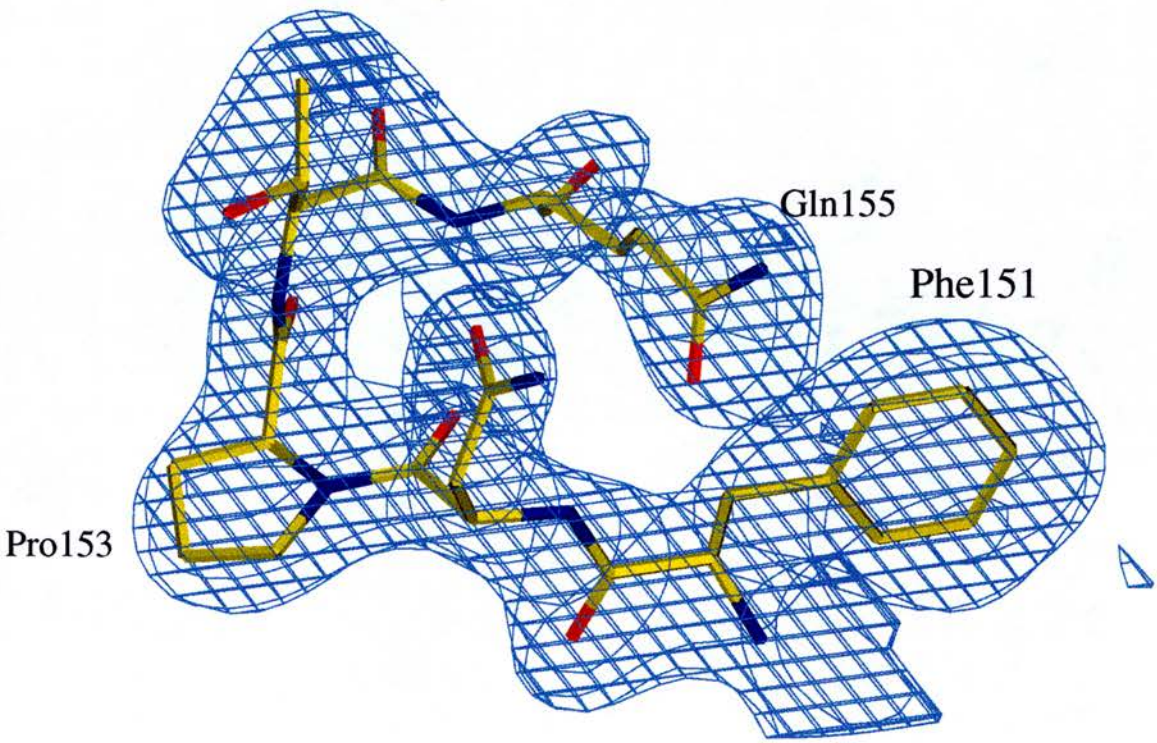
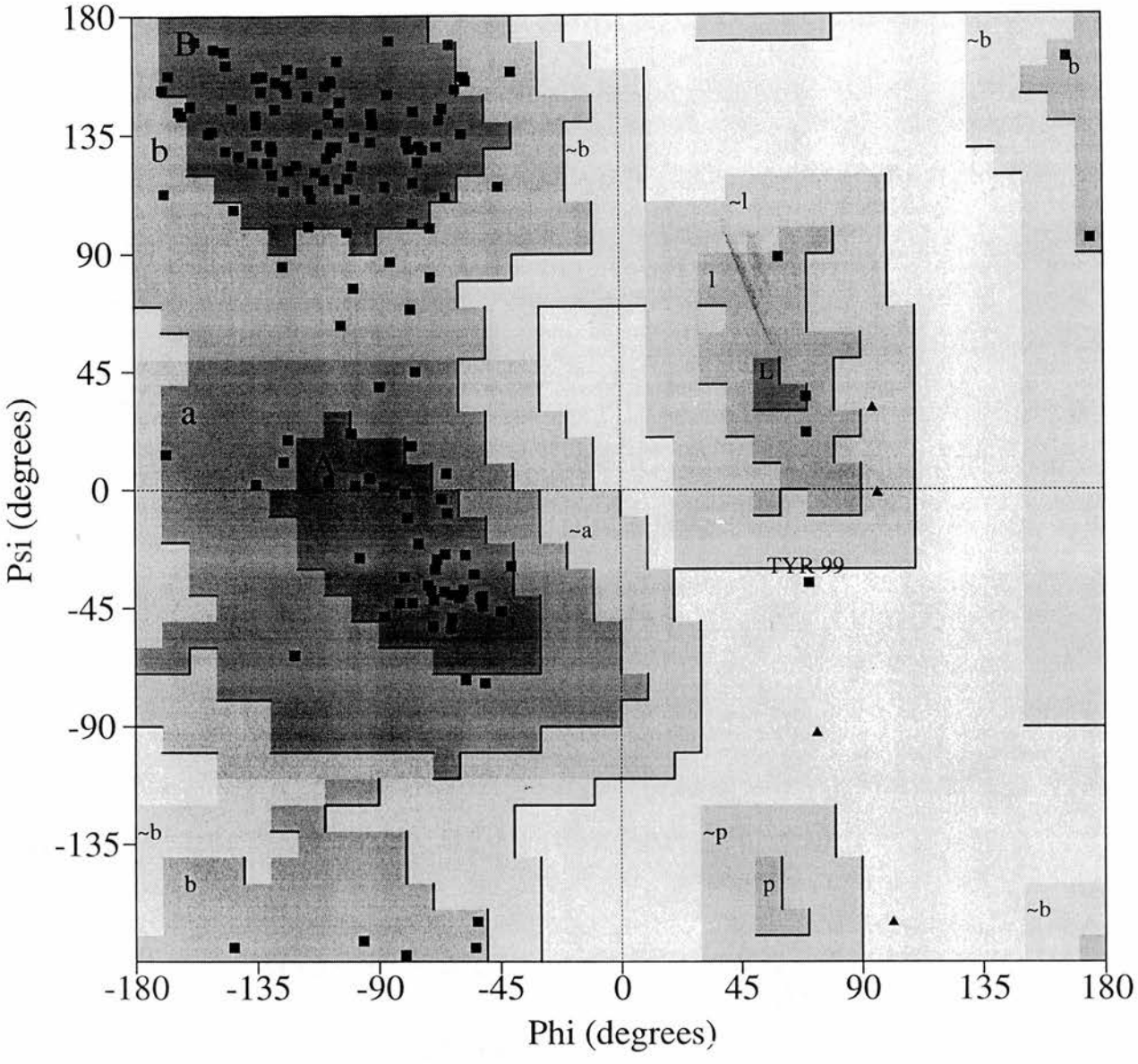


Figure 3-4.The (2Fo-1Fc) electron density map, contoured at 1.5 sigma, of the part of the C-terminus in BLGZ. The map was calculated by SHELXPRO (Sheldrick and Schneider, 1997) after refinement in SHELX and displayed in O (Jones et al., 1991). The density map of this region in BLGZ is much clearer than that of in BLGY (see Figure 3-3).

The solvent content of the crystal is 38.15 %, with the Matthews number 2.0 $Å^3$/Da. There are 8 molecules in the unit cell, with 1 molecule per asymmetric unit. The most complete model comprised 158 residues. Two residues (residue1 and 2) at the N-terminus and two residues (residue161 and 162) at the C-terminus are omitted as there is no density around these regions. The electron density map of the 9 β-strands and the α-helix is very clear. However, as mentioned before, the N-terminus, C-terminus and some loops that connect β-strands have ill-defined density, especially the C-terminus and the CD loop. The AB loop, EF loop and GH loop have rather high B-factors although they were fitted into the density. These disordered loops are commonly found in BLG structures (Brownlow *et al.*, 1997; Qin *et al.*, 1998a). These loops are around the open end of the β-barrel and are flexible, which resulted in difficulty in interpretation of the electron density map. However, it is as a direct result of the flexibility in these regions that the ligand is able to access the calyx without obstruction.

The structure has good geometry as reported by SHELX97 with rms deviation of bond length and angle distance, 0.005Å and 0.018 Å, respectively. The average B-factor for all protein atoms is quite high, 50.0$Å^2$, but is in keeping with the values usually obtained for BLG.

The stereochemical quality of the structure was checked by CCP4 program PROCHECK (Laskowski *et al.*, 1993). The Ramachandran plot is shown in Figure 3-5



Plot statistics

Residues in most favoured regions [A,B,L]	110	76.4%
Residues in additional allowed regions [a,b,l,p]	33	22.9%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	1	0.7%
Number of non-glycine and non-proline residues	144	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	8	
Total number of residues	158	

Figure 3-5. Ramachandran plot of the BLGY structure, produced by PROCHECK (Laskowski *et al.*, 1993). Square symbols represent the non-glycine residues while triangles represent the glycine residues

Essentially, all residues are in the allowed regions, except Tyr99. Tyr99, which adopts a typical γ-turn conformation common to many lipocalin structures with the sequence of TDY, is well defined in the density map and is also in the disallowed region in the lattice X structure (Brownlow *et al.*, 1997), the lattice Z structure (Qin *et al.*, 1998a) and in other lipocalins.

PROCHECK was also used to assess agreement between the main-chain and side-chain parameters of the lattice Y structure with those of well-refined structures at a similar resolution (See Figure 3-6 and Figure 3-7). In Figure 3-6 and Figure 3-7, the lattice Y structure is represented by a solid square, while other reference structures are represented by the dark band. The central line is a least-squares fit to the mean trend as a function of resolution and the width of the band corresponds to a variation of one standard deviation about the mean.

Figure 3-6 shows 6 plots calculated from the BLGY model. Plot (a) presents the quality of Ramachandran plot. This property is indicated by the percentage of the protein's residues that are in the most favoured regions of the Ramachandran plot. Plot (b) shows the peptide bond planarity. This property is measured by calculating the standard deviation of the omega torsion angle. The smaller the value, the closer the ideal value, 180 degrees. Plot (c) presents the bad non-bonded interactions. This property is measured by the number of bad contacts per 100 residues. The bad contact is defined as the distance less than 2.6Å between two atoms. Plot (d) shows the CA tetrahedral distortion. This property is measured by calculating the standard deviation of the "Zeta torsion angle", which is defined by the following 4 atoms: CA, N,C,CB. Plot (e) is the standard deviation of the main-chain hydrogen bond energies calculated using the method of Kabsch and Sander (1983). Plot (f) is the overall G-factor. The G-factor is a measure of the normality of the stereochemical parameters, like bond length, bond angle, torsion angle, compared to the other structures. The overall G-factor is calculated from an average of the different G-factors for each residue in the BLGY model.

The structure was of acceptable quality as the stereochemistry is in good agreement with what is expected for a structure at this resolution.

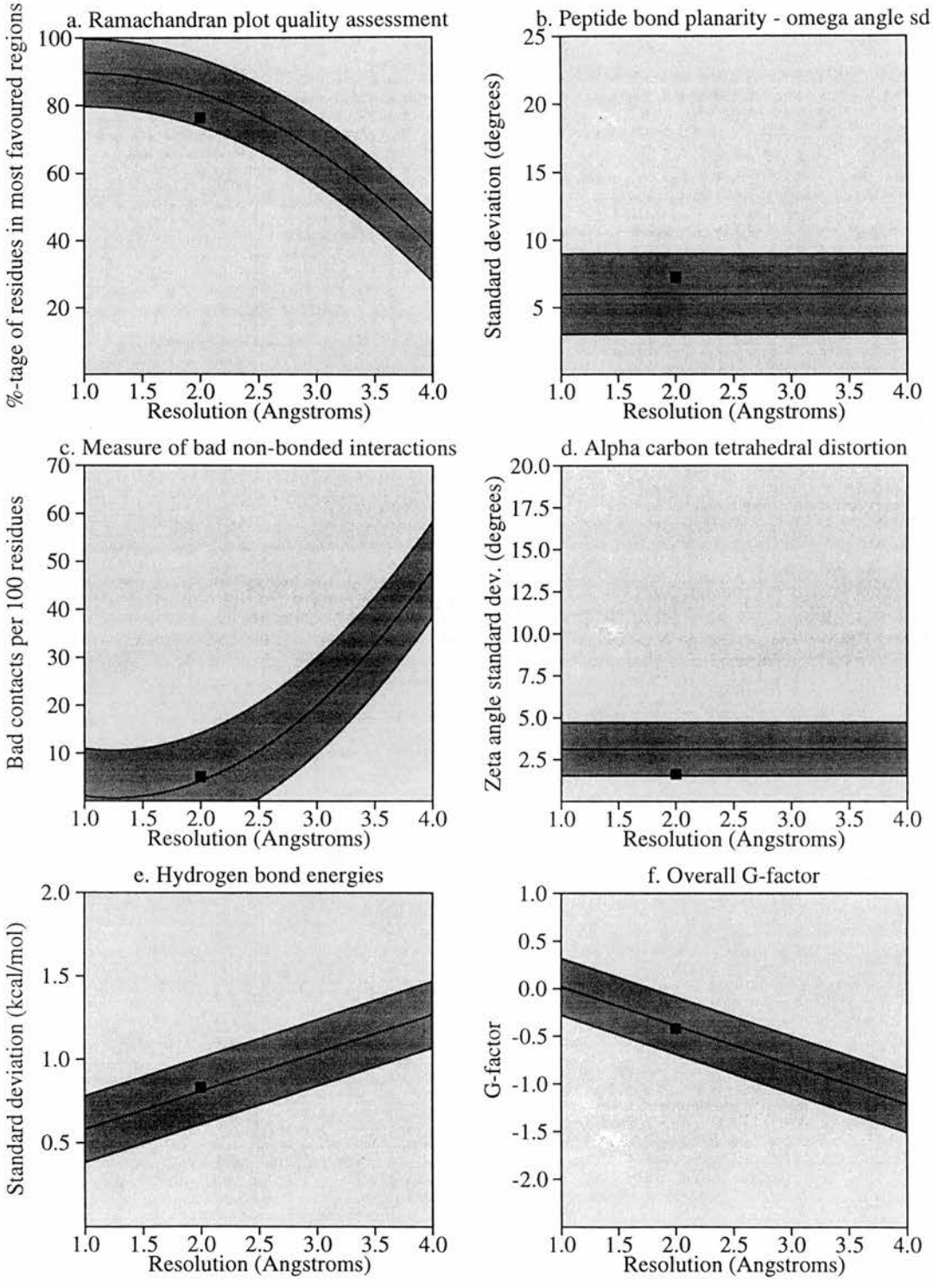


Figure 3-6. Plots of the comparison of the main chain parameters of BLGY with those of other structures solved to a similar resolution, produced by the program PROCHECK (Laskowski *et al.*, 1993).

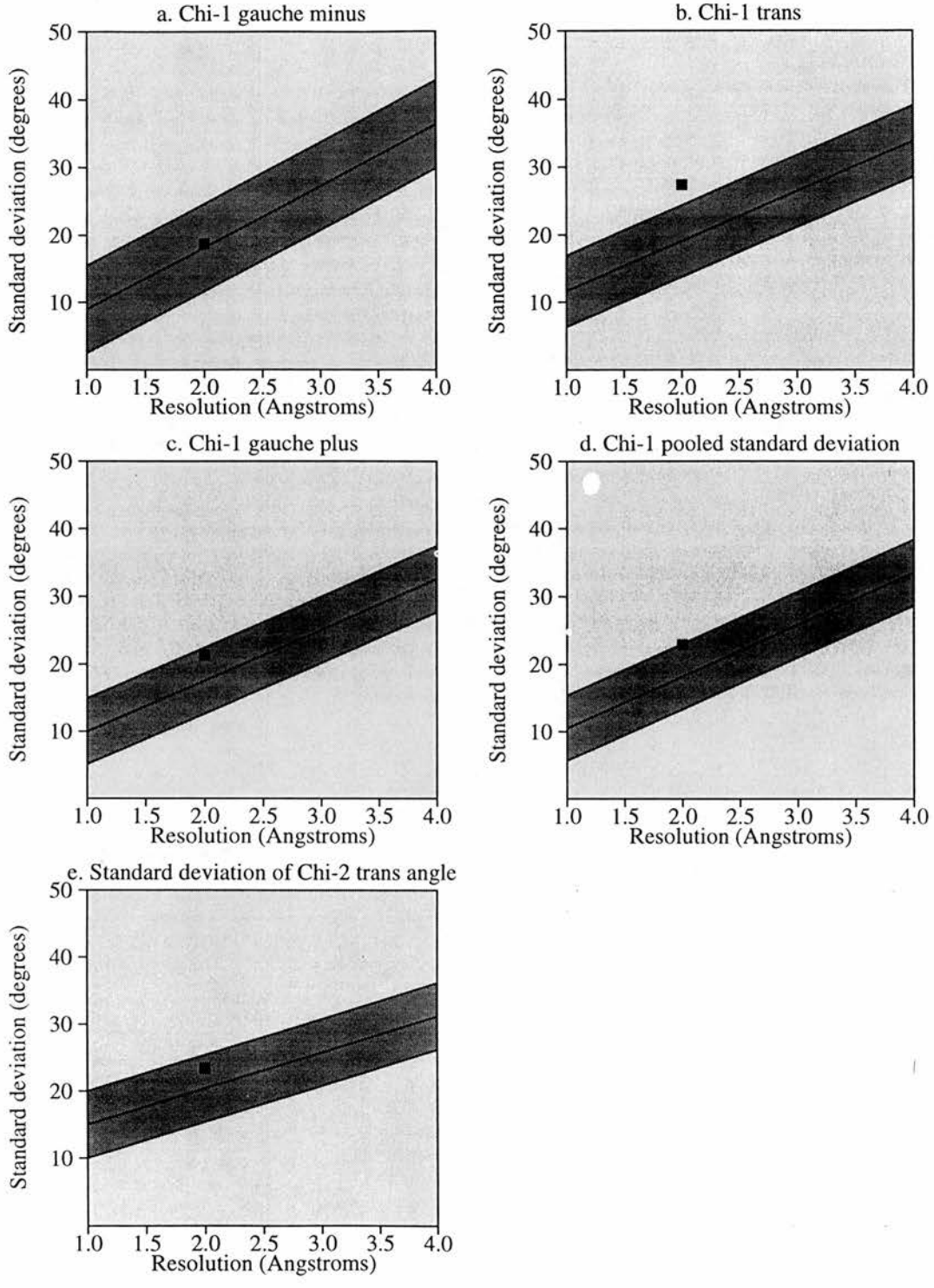


Figure 3-7. Plots of the comparison of the side chain parameters of BLGY with those of other structures solved to a similar resolution, produced by the program PROCHECK(Laskowski *et al.*, 1993).

3.2 Overall structure

The overall structure of BLG in lattice Y form (BLGY) is very similar to the structures of two other crystal forms, lattice X (BLGX) and lattice Z (BLGZ) as shown in Figure 1-1.

The core region of BLGY is identical to that of BLGX and BLGZ. However, the N-terminus, the C-terminus and some loops that connect the β-strands adopt different conformations in these three crystal forms, especially the AB, CD, EF and GH loops. These loops are highly flexible and rather disordered as mentioned in sections 3.1. Similar to the BLGX and BLGZ structures, the BLGY structure also contains two disulphide bridges. Since the CD loop and C-terminus in the lattice Y structure are very disordered and the density maps around these two regions are rather poor, the disulphide bridge, 66-160, appears mobile and the density of this disulphide bridge is very weak.

BLG exists as a dimer in physiological conditions and at room temperature. In the lattice Y structure, the asymmetric unit in the crystal contains only a single monomer and the dimer is created by a crystallographic 2-fold axis. The dimer therefore has exact 2-fold symmetry. Strand I and AB loop form the dimer interface. The interactions that occur at the dimer interface of BLGY are similar to those of BLGX (Brownlow *et al.*, 1997) with the exception of the absence of one hydrogen bond (N/Ala34 →OD1/Asp33) in the BLGY structure. However, the hydrogen bonds between the adjacent AB loops at the dimer interface are weaker than these in BLGX and BLGZ structures because of longer separation between the adjacent AB loops. This movement of the AB loop will be discussed in section 3.3.

One end of the β-barrel is open to ligand binding as the EF lifts up to open up the entrance to the calyx in the same manner as in the complexed structures of palmitate-bound BLG, which will be discussed in chapter 4. This important feature of the EF loop is discussed further in section 4.5. Furthermore, the flexibility of the CD, EF, GH loops at the open end of the β-barrel allows the ligand to enter the calyx without hindrance.

3.3 Comparison of BLGY and BLGZ structure

In sections 3.3 and 3.4, the BLGY structure will be compared with the BLGZ and BLGX structures, respectively. The most complete BLGY structure contains residues 3 to 160. However, as mentioned in section 3.1.4, the density maps of the CD loop and the C-terminus are very poor and these two regions are not well-fitted into the density. Therefore, the difference of the CD loop and C-terminus in the BLGY structure compared to the BLGZ and BLGX structures remains questionable as these two regions in the BLGY structure have not yet been proven to be correct.

The superimposition of the structures of BLG in the lattice Y (BLGY) and the lattice Z (BLGZ) forms is shown in Figure 3-8.

The structures of BLGY (at pH 7.5) and BLGZ (pH 7.1; PDB entry 1BSY) are very similar, the exception being the GH, AB, CD loops, the loop before the α-helix and C-terminus. The rms between corresponding backbone atoms of 116 residues in BLGY and BLGZ structures is 0.737Å with the GH, AB, CD loops, the loop before the α-helix and C-terminus excluded.

The GH loop, a hairpin loop held by the strand G and H, is at the open end of the calyx and is very flexible with high B-factors seen in both structures. Comparison of the BLGY structure with the BLGZ structure shows that the divergence of the GH loop begins at Asn109 and ends at Ser116. Moreover, this loop in BLGY moves toward the entrance of the binding site with the maximum shift of 5.0Å at the CA atom of Glu115. However, it seems that this movement has no effect on the access of ligands, at least for fatty acid, as there is still enough space to allow the ligands to entry. The AB loop is located at the dimer interface. In BLGY, this loop moves slightly toward the mouth of the calyx. This movement results in the adjacent AB loops in the dimer interface moving further apart. Therefore, the hydrogen bonds between these two adjacent loops are weaker as the bond distances are longer. This weakening of the hydrogen bonds is probably related to the observation that the dimer dissociates into monomer more readily as pH increases (Georges and Guinand, 1960). However, it is highly possible the different relative disposition of the monomer in these two crystal forms is a result of the different crystal packing forces.

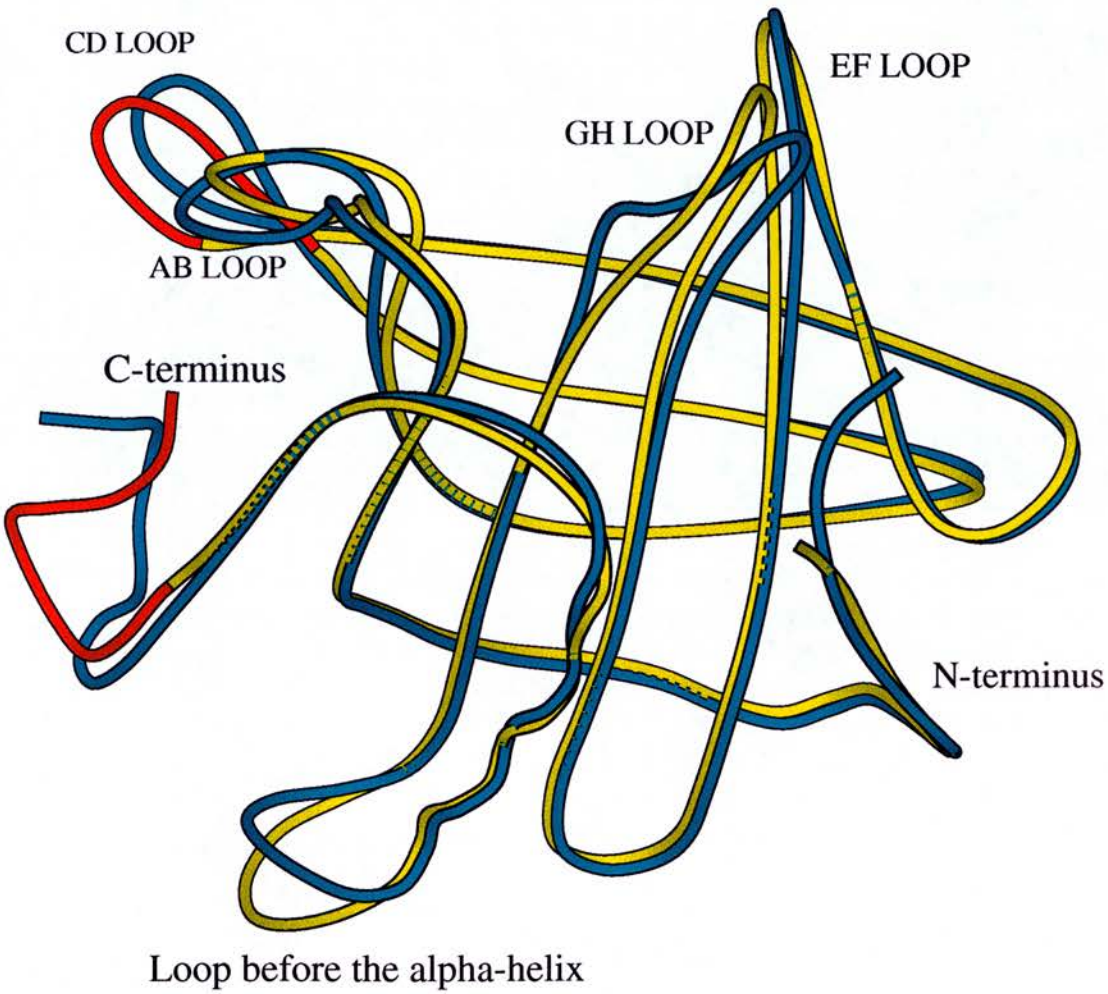


Figure 3-8.Diagram of the superimposition of the BLGY (yellow) and BLGZ (cyan) structures, prepared by MOLSCRIPT (Kraulis, 1991). By comparison, the differences of these two structures occur at the AB, GH loops and the loop before the alpha-helix. The problematic regions of BLGY, the CD loop and the C-terminus, are shown in red.

The other divergent loop is the loop before the α-helix. Comparison of these two structures reveals divergence begins at Pro126 and ends at Asp130 with the maximum shift, 3.87Å, at the CA atom of Val128. This loop is well-defined in the density map and is rather rigid in the BLGY structure. The movement of this loop is probably a consequence of the different crystal packing force. In the BLGY structure, the other symmetry related molecule is nearby and the shift of this loop makes better hydrophobic contacts and hydrogen bond contacts with the other symmetry related molecule.

There are two short 3_{10} helix turns at the C-terminus of the BLGZ which are absent in the BLGY structure. However, as mentioned before, the structure of the C-terminus in the BLGY structure is doubtful. Therefore, the absence of these two 3_{10} helical turns cannot be taken as definite.

3.4 Comparison of BLGY and BLGX structure

The superimposition of the structures of BLG in the lattice Y (BLGY) and the lattice X (BLGX) crystal forms (Brownlow *et al.*, 1997) is shown in Figure 3-9. The rms between corresponding backbone atoms of 113 residues in BLGY and BLGX structures is 0.692Å with the EF, GH, AB, CD loops, the loop before the α-helix and C-terminus excluded.

Comparison of BLGY (pH 7.5) and BLGX (pH6.5) reveals the most distinctive difference is in the movement of the EF loop. The EF loop is folding toward the calyx in the BLGX structure whereas the EF loop in the BLGY structure is folding back to open up the entrance to the calyx. Since lattice X and lattice Y are crystallised at the two ends of the Tanford transition (Tanford *et al.*, 1959), the BLGY structure confirmed the important role of the EF loop in the Tanford transition as discussed in the section 1.1.7. Moreover, in the BLGY structure, Glu89 buried in the BLGX structure at pH 6.5 is exposed in the BLGY structure along with the movement of the EF loop. This is consistent with the observation in the BLGZ structure at pH 7.1 (Qin *et al.*, 1998a) and the structure of palmitate-bound BLG, which will be discussed in chapter 4.

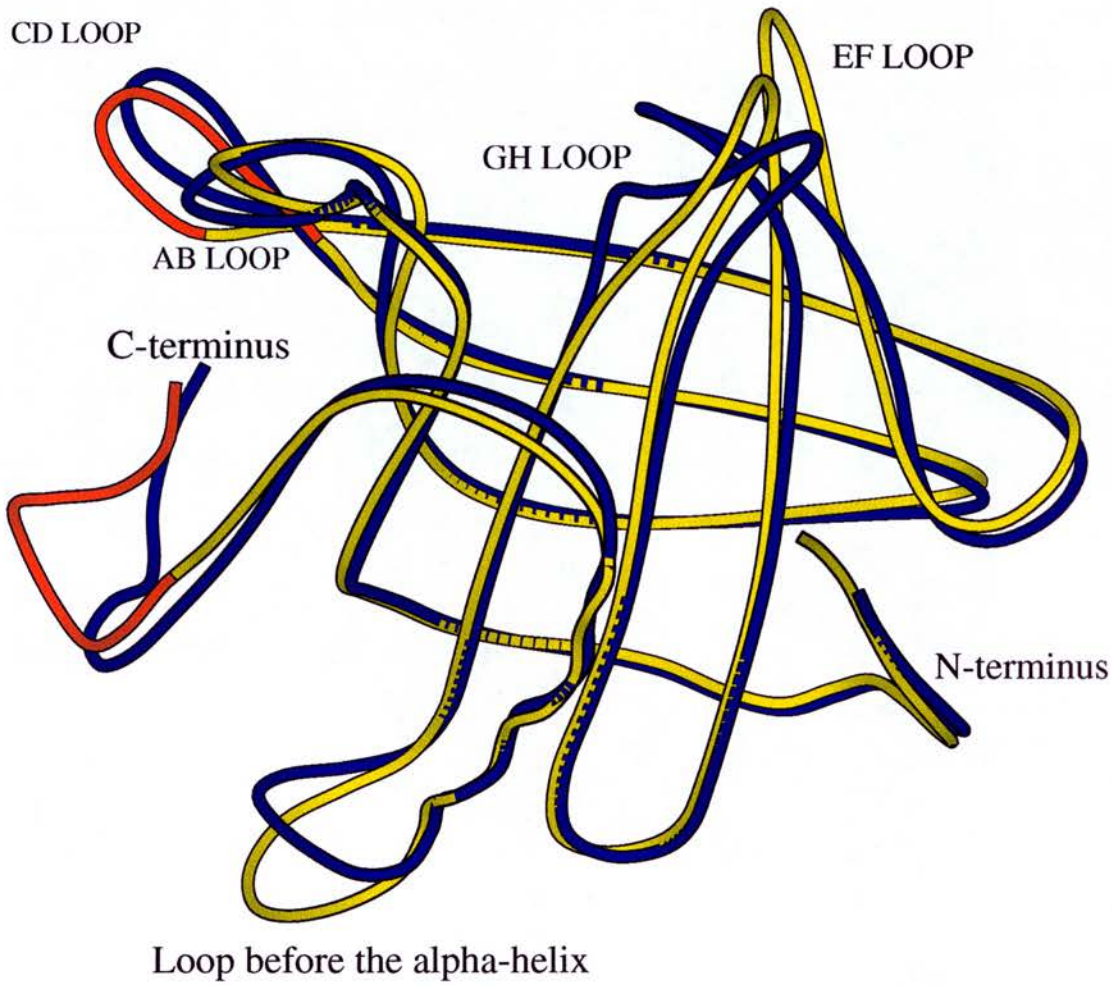


Figure 3-9.Diagram of the superimposition of the BLGY (yellow) and BLGX (blue) structures, prepared by MOLSCRIPT (Kraulis, 1991). By comparison, the differences of these two structures occur at the AB, EF, GH loops and the loop before the alpha-helix as labelled. The problematic regions of BLGY, the CD loop and the C-terminus, are shown in red.

Recently, it has been reported (Qin *et al.*, 1999) that the EF loop adopts different conformations at pH 7.1 in the structures of different genetic variants. In BLGA structure (pH 7.1), the EF loop adopts the open conformation whereas in BLGB structure (pH 7.1), this loop adopts the close conformation, the same as the structures in the lattice X form (Brownlow *et al.*, 1997) and in the lattice Z form (Qin *et al.*, 1998a) at pH about 6.5. However, it has been noted by the author of this paper (Qin *et al.*, 1999) that it is not clear whether the conformational difference of the EF loop is actually due to different generic variants or other uncertain reasons. This observation of the alternative conformation in the EF loop at pH 7.1 is contradiction with the result discussed above where the BLGY structure was obtained from the B variant (pH 7.5) and the EF loop was in the open conformation. Since the conclusion of the conformational difference of the EF loop because of different genetic variants can not be taken as definite, the reasons resulting in the contradiction between two structures require further investigation.

The other divergent loops are the AB, GH ones and the loop before the α-helix. The differences in these loops are also found between the BLGY and BLGZ structures as discussed in section 3.3. It seems these loops are located in a different position in the BLGY structure while these loops are in the same position in BLGX and BLGZ structures. It is therefore proposed the dispositions of these loops are due to the different crystal packing force instead of the influence of pH as BLG X is crystallised at pH below 7.0 while BLGX and BLGY are crystallised at pH above 7.0.

3.5 Discussion

The refinement of the BLG structure in lattice Y form has been a problem for many years. Many studies have been focused on the lattice Y structure (Papiz *et al.*, 1986; Hambling, 1990; McAlpine. 1991; Bewley et al., 1997) since the first structure was solved in 1979 by Green (Green *et al.*, 1979). However, successful refinement of BLGY structure has not really been achieved although a variety of methods have been tried. The problems are either the disagreement between the model and the data or the poor geometry of the refined model.

Each of these data sets was collected independently under different system and crystals were crystallised under different conditions. This maybe the cause of some of the problems encountered in the refinement.

Although the core region of these models was identical, the loops varied from model to model. The most problematic regions occur in the CD loop and the C-terminus. The C-terminus, which is well-defined in the lattice Z form, has rather weak density and is poorly defined in the lattice Y structure. The much greater flexibility of the C-terminus in BLGY is probably the consequence of the weaker crystal packing force in the lattice Y form. The CD loop is commonly disordered in all BLG structures published. These two regions probably have some effects in hindering refinement. Although a variety of approaches have been tried, the successful refinement still can not be achieved. For example, the residues from 60 to 66 (CD loop) and 152 to 160 (C-terminus) were firstly removed and this partial model was then subjected to refinement. However, (Fo-Fc) map calculated after refinement fails to show continuous main chain density in these disordered regions. Secondly, the re-built model subjected to simulated annealing has proven no more successful in decreasing the R-factor and after stimulated annealing, the model was out of the density. In addition, the possibility of using the structure of C-terminus in the lattice Z form to locate the same area in lattice Y form was considered but failed as the main chain does not fit into the density properly and so reflects a clear difference in these two crystal forms at this point. Moreover, the rebuilding of these disordered regions fails to reduce the R-factor and some residues were being pulled back toward the position that they had occupied before the rebuilding.

The other possibility is that the problem may lie in the quality of the X-ray data. The data were assessed carefully using DENZO. As summarised in Table 2-F, 131207 reflections between 23.0 to 2.0Å resolution were measured and the completeness was 99.7% with the multiplicity of 12.68. The overall $I/\sigma(I)$ is 27.22 while the $I/\sigma(I)$ of the outer shell (2.03Å to 2.00Å) is 6.30. The high $I/\sigma(I)$ indicates the reflections are strong even at high resolution. The overall R_{merge} is 5.1% while the R_{merge} of the outer shell (2.03Å to 2.00Å) is 20.8%. The low R_{merge} shows the intensities of symmetry-related reflections are in a good agreement of each other.

The overloading problem has also been considered. Since the data set was collected at the Daresbury synchrotron, there were a few reflections (approximately 20 reflections) at low resolution that were overloaded due to strong beam intensity and long exposure time needed for the collection of the high resolution data. After the high resolution data set was collected, the detector was moved further away the crystal to collect the low resolution data. Careful data collection strategies, such as decreasing the exposure time, were taken to avoid the overloading of any reflection. The high resolution data with the overloaded reflections excluded and the low resolution data were then merged together. The final R_{merge} was 5.1%. The low merging R-factor shows these two data sets scaled well. This merged data set was used for the refinement is this thesis. The statistics, including low R_{merge}(5.1%), high $I/\sigma(I)$ (27.22) and high completeness (99.7%), show the quality of the data is good. Therefore, it would appear that the difficulty of refinement does not come from the quality of the data.

Moreover, the possibility of the alternative space group with lower symmetry has also been examined. The same data set was processed in the space group C2. The date processing results are as follows: a=79.73Å, b=55.18Å, c=66.68Å, $\alpha=\gamma=90°$, $\beta=89.99°$; the completeness was 99.0% with the multiplicity of 6.99. After data processing, molecular replacement with the program AMORE was performed. In the space group $C222_1$, there is a monomer in an asymmetric unit while there is a dimer in an asymmetric unit in the space group C2. In other words, the dimer of BLG is a crystallographic dimer in $C222_1$ while it is non-crystallographic dimer in C2. Therefore, a dimeric search model was used for molecular replacement in the space group C2. The dimeric model, generated in O, consists of two symmetry equivalent molecules (monomers) which are related by a crystallographic 2-fold axis along the b axis (the molecular dyad) in the space group $C222_1$. After rotation and translation functions with AMORE, a solution ($\alpha=0.00$, $\beta=180.00$, $\gamma=-90.00$, Tx=0.3452, Ty=0.0000, Tz=0.2500) with a high correlation coefficient (70.7%) and a low R-factor (33.2%) was given. Following molecular replacement, the dimeric model was subject to refinement with SHELX97 by performing the same procedures as mentioned in section 3.1.4. However, the refinement result did not achieve any more

success as the R-factor is 27.8%, which was worse than that obtained from the space group $C222_1$. Thus, it confirms that the space group $C222_1$ is correct.

An other approach to the problem was to collect data from another crystal grown in a different plate and different drop to investigate if the crystal quality has effects on the refinement. This data set was collected using home source. 77480 reflections between 20.0 to 2.0Å resolution were measured and the completeness was 99.0% with the multiplicity of 7.46. The overall $I/\sigma(I)$ is 16.00 while the $I/\sigma(I)$ of outer shell (2.03Å to 2.00Å) is 2.43. The overall R_{merge} was 5.4% while the R_{merge} of outer shell (2.03Å to 2.00Å) is 32.7%. These data were refined against the model with the same procedure as described in section 3.1.4. However, the refinement did not show any improvement since the R-factor fluctuated around 26%, the same as the previous data set.

In conclusion, the refinement problem probably does not lie in the quality of the data set and the crystal quality. It is therefore tempting to speculate that the unsuccessful refinement was due to the difficulty in building the disordered regions, such as C-terminal and the CD loop, although the actual reasons resulting in hindering the refinement require further investigation.

4. Structure determination of BLGZ complexed with palmitate

This chapter will describe in four parts the structural studies of BLG complexed with palmitate. The first part is to determine and refine the complexed structure (section 4.1). The second part is to examine the complexed structure and the interactions between palmitate and BLG (sections 4.2 to 4.4). The third part is to compare the structure of palmitate-bound BLG with the other structures (sections 4.5 to 4.8). Finally, the possibility of the existence of the second binding site will be discussed (section 4.9).

4.1 Molecular replacement

4.1.1 Molecular replacement using a monomer of lattice X

The structure was solved by molecular replacement using AMORE (Navaza and Saludjian, 1997). Data within the resolution range 10-4Å and an integration radius of 18 Å were used to calculate the rotation and translation functions.

A monomer of the crystal structure of dimeric BLG refined at 1.8Å resolution in triclinic Lattice X (space group P1; a=37.8Å, b=49.6Å, c=56.6 Å, α=123.4°, β=97.3°, γ=103.7°; PDB ID 1BEB) was used as the search model (Brownlow *et al.*, 1997) in the structure determination of the Lattice Z crystal forms containing the palmitate. Although the 3.0Å lattice Z structure (Brownlow *et al.*, 1997) could have been used, the higher resolution lattice X structure was preferred as the best starting model.

The input values for variables used with the AMORE program are listed in Table 4-A.

Table 4-A. Input values for variables used with the AMORE program

Resolution (Å)	10-4
P1 box size (Å,°)	(75, 75, 75, 90, 90, 90)
Radius of integration (Å)	18
Angular step (°)	1.0

The top six peaks of the rotation function are listed in Table 4-B.

Table 4-B. The top six peaks of the rotation function using the lattice X monomer as the search model

Peak	(α, β, γ,)	Correlation coefficient (%)	Sigma (σ)
1	(78.12, 26.87, 206.14)	22.2	4.86
2	(41.88, 153.13, 26.14)	22.2	4.86
3	(95.93, 140.45, 121.85)	16.9	3.70
4	(24.07, 39.55, 301.85)	16.9	3.70
5	(92.25, 65.18, 96.91)	16.5	3.16
6	(27.75, 114.82, 276.91)	16.5	3.16

After checking with the CCP4 program ROTMAN (Dodson,1997), it became apparent that peaks 1 and 2 are symmetry equivalent.

The maximum peak (4.86σ) in the rotation function and next highest peak (3.70σ) were used to calculate the translation function. After the translation search, the maximum peak (α=78.12 β=26.87 γ=206.14) gave a distinct solution with a high correlation coefficient of 60.8% and a low R-factor of 38.9%. The next solution has a lower correlation coefficient of 34.2% and a higher R-factor of 49.8%.

The top five solutions of translation function using this maximum peak are listed in Table 4-C.

Table 4-C. The top five solutions of translation function using the maximum peak obtained from rotation function (α=78.12 β=26.87 γ=206.14)

Solutions of translation function	Fractional coordinates (Tx, Ty, Tz)	Correlation coefficient (%)	R-factor (%)
1	(0.3317, 0.2440, 0.3664)	60.8	38.9
2	(0.6639, 0.9092, 0.2827)	34.2	49.8
3	(0.9969, 0.5724, 0.3675)	32.6	50.7
4	(0.3338, 0.2454, 0.1586)	32.5	50.1
5	(0.6617, 0.9109, 0.4085)	32.2	51.4

The packing of the molecules in the unit cell was checked by the program PACMAN (Kleywegt *et al.*, 1996) and also displayed in O (See Figure 4-1). Both programs

showed that the packing of the first solution was reasonable with no symmetry-related molecules clashing into each other.

It was obvious that the first solution was correct since it had a significantly higher sigma value, higher correlation coefficient and lower R-factor when compared to the second solution. The packing of this solution was also reasonable. As well as peak 1, peak 3 (α=95.93 β=140.45 γ=121.85) obtained from the rotation function was used to calculate the translation function. However, it failed to give a distinct solution like peak 1. In addition, the solutions that it did give had much lower correlation coefficients (approximately 20.0%), and much higher R-factors (approximately 55.5%).

The space group $P3_221$ was also confirmed by the translation function. The rotation function in space group $P3_121$ give the same solutions as in space group $P3_221$. However, when the translation function was calculated using peak 1 (α=78.12 β=26.87 γ=206.14) of the rotation function, no distinct solution was found in space group $P3_121$. The top five solutions of translation function in space group $P3_121$ using peak 1 are listed in Table 4-D.

Table 4-D The top five solutions of translation function in space group $P3_121$ using the peak 1 obtained from rotation function (α=78.12 β=26.87 γ=206.14)

Solutions of translation function	Fractional coordinates (Tx, Ty, Tz)	Correlation coefficient (%)	R-factor (%)
1	(0.3862, 0.2952, 0.3666)	26.8	53.1
2	(0.9620, 0.8709, 0.3658)	26.7	52.5
3	(0.7393, 0.6586, 0.3669)	26.4	52.6
4	(0.3383, 0.6133, 0.2000)	26.0	53.2
5	(0.0209, 0.9375, 0.3671)	25.7	53.4

The best solution in $P3_121$ gave a lower correlation coefficient of 26.8% and a higher R-factor of 53.1% whereas the best solution in $P3_221$ gave a correlation coefficient of 60.8% and an R-factor of 38.9%.

Following the translation function, rigid-body refinement of the optimum molecular replacement solution was performed. This reduced the R-factor slightly from 38.9% to 38.0% and the Eulerian angles and fractional coordinates shift slightly to α=78.28 β=26.50 γ=206.33 Tx=0.330 Ty=0.246 Tz=0.3663.

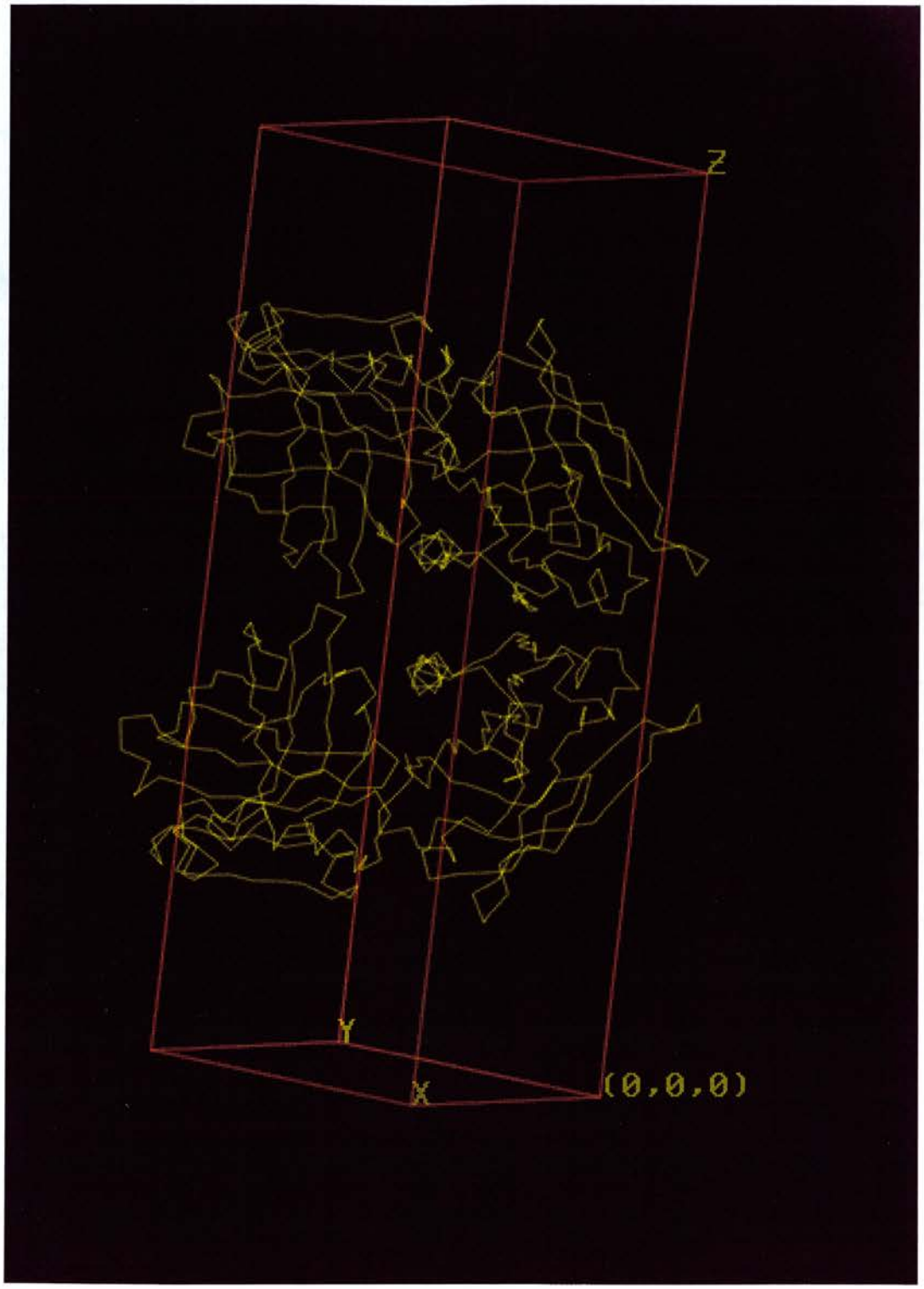


Figure 4–1.Packing diagram of the molecular replacement solution, prepared by O (Jones et al., 1991).

4.1.2 Molecular replacement using lattice Y model

The BLG Lattice Y structure (space group $C222_1$; a=55.5Å, b=82.1Å, c=66.8, $\alpha=\beta=\gamma=90°$) was also used as a search model in the structure determination of the lattice Z crystal form containing palmitate. Following the same procedure as in section 4.1.1 and using the same input variables as were applied to the lattice X monomer, the rotation function gives the maximum peak (α=97.74, β=47.43 γ=352.49) of 5.41σ and the next highest peak (α=96.09, β=12.78 γ=136.42) of 3.94σ. Both were used to calculate the translation function. After the translation search, the maximum peak (α=97.74, β=47.43 γ=352.49) gave the best solution with a correlation coefficient of 38.3% and an R-factor of 48.6% while the next solution had a correlation coefficient of 28.0% and an R-factor of 52.3%. The top five solutions of translation function are listed in Table 4-E.

Table 4-E. The top five solutions of translation function using the maximum peak obtained from rotation function (α=97.74, β=47.43 γ=352.49)

Solutions of translation function	Fractional coordinates (Tx, Ty, Tz)	Correlation coefficient (%)	R-factor (%)
1	(0.6686, 0.9090, 0.3007)	38.3	48.6
2	(0.6697, 0.9090, 0.2582)	28.0	52.3
3	(0.6679, 0.9093, 0.0731)	27.2	53.2
4	(0.3371, 0.2411, 0.3425)	26.6	53.0
5	(0.6714, 0.9107, 0.1160)	26.3	53.4

The packing of the molecules in the unit cell was checked by the program PACMAN (Kleywegt *et al.*, 1996) and also displayed in O. Both programs showed that the packing of the first solution was reasonable with no symmetry-related molecules clashing into each other.

The first solution was correct since it had a higher sigma value, higher correlation coefficient and lower R-factor when compared to the second solution. The packing of this solution was also reasonable.

The next highest peak of rotation function (α=96.09, β=12.78 γ=136.42) gave no significant solution and all solutions are with a lower correlation coefficient around 17.0% and a higher R-factor around 56.0%.

The solutions obtained using lattice Y as the search model were generally poor and therefore the solution obtained from molecular replacement using lattice X as the search model was used for further refinement and discussed in this thesis.

4.1.3 Refinement and model building

Following rigid-body refinement, the model was subjected to simulated annealing using data between 10Å and 2.5Å. The slow cooling protocol for refinement by simulated annealing was used (Brunger *et al.*, 1990b). Firstly, the model was heated to 3000K, followed by 50 steps of molecular dynamics integration performed with a time step of 0.5fs. After each molecular dynamics calculation, the temperature was decreased by 25°K to the final temperature of 250°K. The R-factor dropped to 34% (R_{free}=38.1%) after simulated annealing. The electron density maps, (3Fo-2Fc) and (Fo-Fc), calculated in X-PLOR and converted in MAPPAGE (Jones, 1992) were displayed in O. In the (3Fo-2Fc) map (see Figure 4-2), some density appeared in the proposed binding site inside the calyx (Cho *et al.*, 1994).

Although the electron density was not clear enough to fit the entire palmitate, it looked promising as the shape of the density is elongated, like an extended carbon chain. Actually, this part of the density was confirmed as the tail of the palmitate at the final refinement stage. The density of the head of the palmitate was missing in the (3Fo-2Fc) map. However, in the (Fo-Fc) map (see Figure 4-3), some density appeared around the mouth of the calyx near to the CD loop.

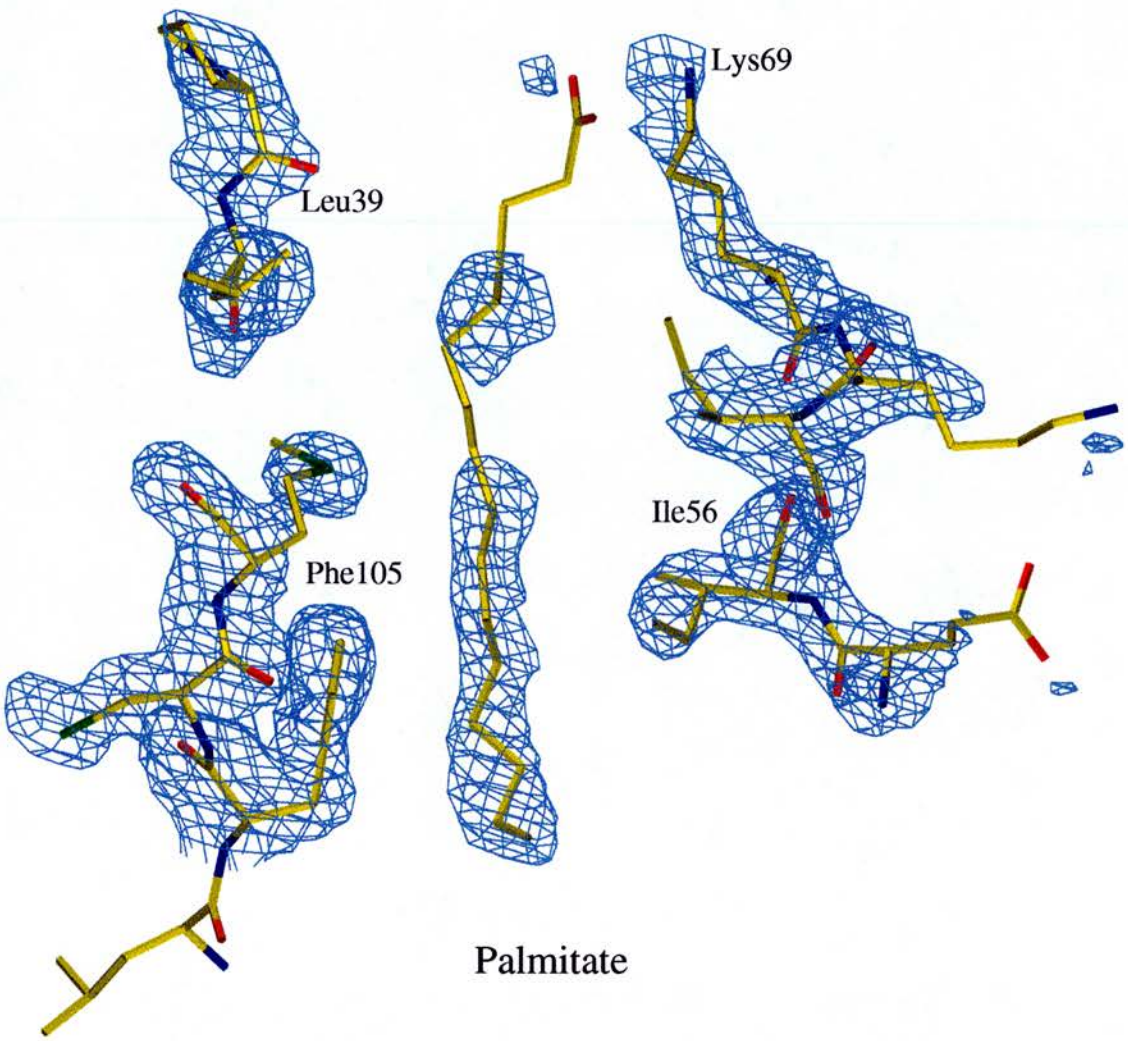


Figure 4-2.The (3Fo-2Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 1.2 sigma, was calculated by XPLOR (Brunger, 1992) after simulated annealing refinement and displayed in O (Jones et al., 1991). The density of tail of the palmitate appears in the binding site.

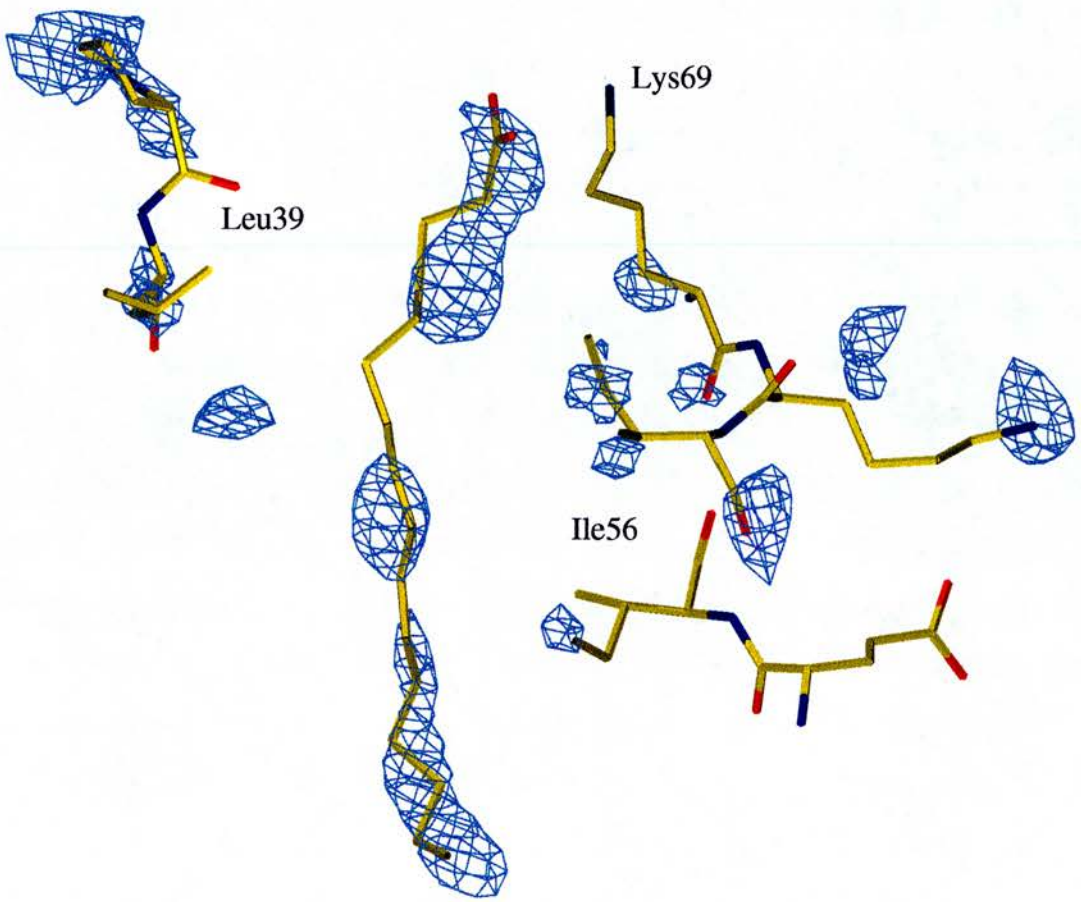


Figure 4-3.The (Fo-Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 2.0 sigma, was calculated by XPLOR (Brunger, 1992) after simulated annealing refinement and displayed in O (Jones et al., 1991).

The density was very weak, therefore little could be said about the location and orientation of palmitate at this stage. Following simulated annealing, 600 cycles of positional refinement were carried out. Individual B-factors were also refined after positional refinement. This resulted in a decrease of the R-factor from 34% to 29.8% (R_{free} from 38.1% to 33.8%). However, the density of palmitate was still not significantly improved. Following individual B-factor refinement, H_2O molecules were added using the program WATERPICK supplied in X-PLOR. The positions of H_2O were checked manually in O firstly to ensure the location was appropriate and secondly that the H_2O made a reasonable hydrogen bond contact with the protein. The R-factor decreased to 25.2% (R_{free}=29.0%).

From the density map (see Figure 4-4), it was obvious the EF loop adapts the alternative conformation compared to the original model (BLGX model). Therefore, the decision was taken to re-build manually the EF loop in order to clarify this region. The density in this region was weak, however careful inspection allowed tracing of the main chain through continuous density. As the (3Fo-2Fc) map (see Figure 4-4) was more clear than the (Fo-Fc) map, it was used to re-build EF loop. Following rebuilding, further positional and B-factor refinement resulted in a reduction of the R-factor to 24.4% (R_{free}=28.8%) and enhanced clarity in the density map around the bound palmitate (see Figure 4-5).

The broken density seen in Figure 4-2 had joined together and stronger density appeared around the open end of the calyx. The rebuilt model was subjected to further refinement with SHELX97-SHELXL. After 5 cycles of positional and B-factor refinement, more H_2O molecules were added using SHELXWAT. A (2Fo-1Fc) map (see Figure 4-6) was calculated using SHELXPRO and displayed with O. The refinement and addition of H_2O in SHELXL improved the density sufficiently to allow building of palmitate, resulting in a decrease of R-factor to 21.5% (R_{free}=25.2%).

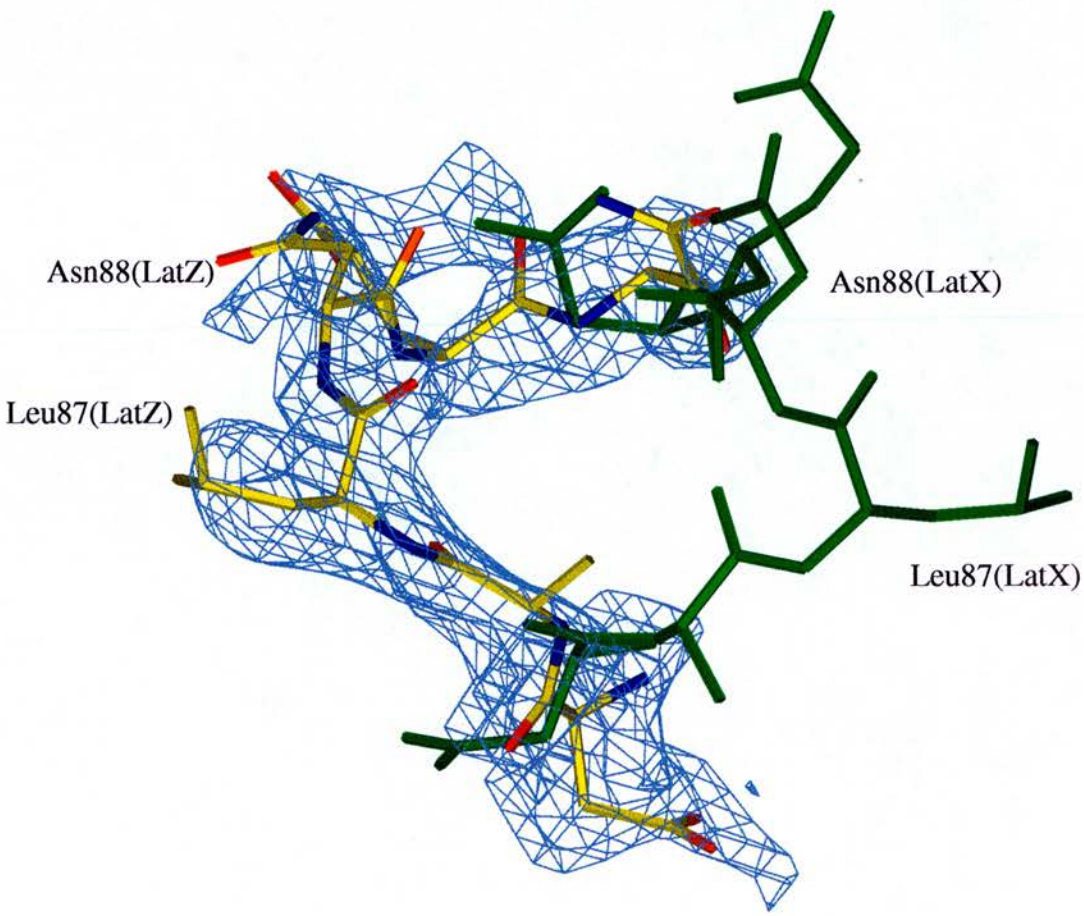


Figure 4-4.The (3Fo-2Fc) electron density map of the EF loop with the refined BLGZ structure superimposed. The EF loop of the original model (BLGX) is shown in green. The map, contoured at 0.8 sigma, was calculated by XPLOR (Brunger, 1992) after the addition of the water molecules and displayed in O (Jones et al., 1991).

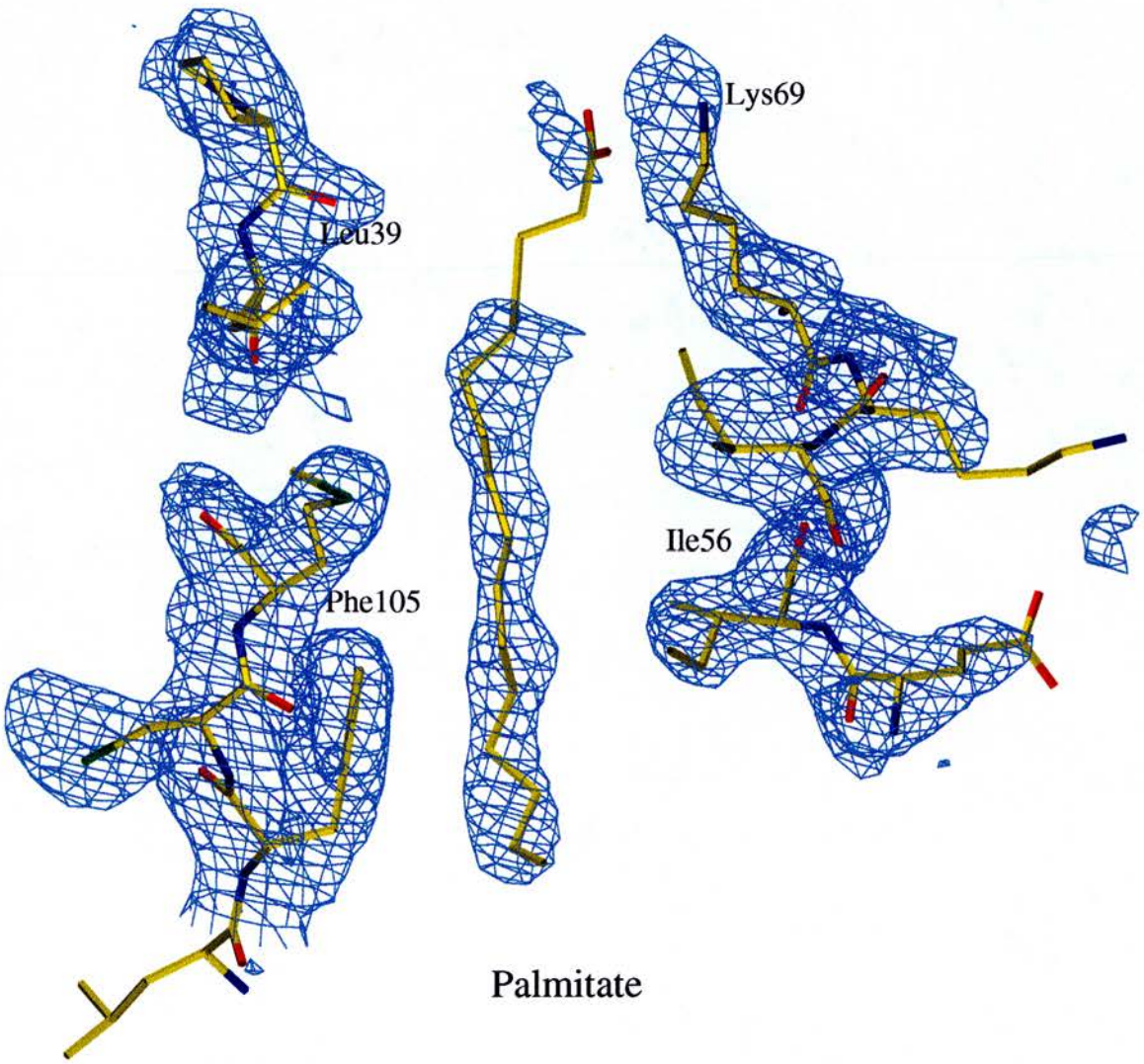


Figure 4-5.The (3Fo-2Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 1.0 sigma, was calculated by XPLOR (Brunger, 1992) after rebuilding the EF loop and displayed in O (Jones et al., 1991).

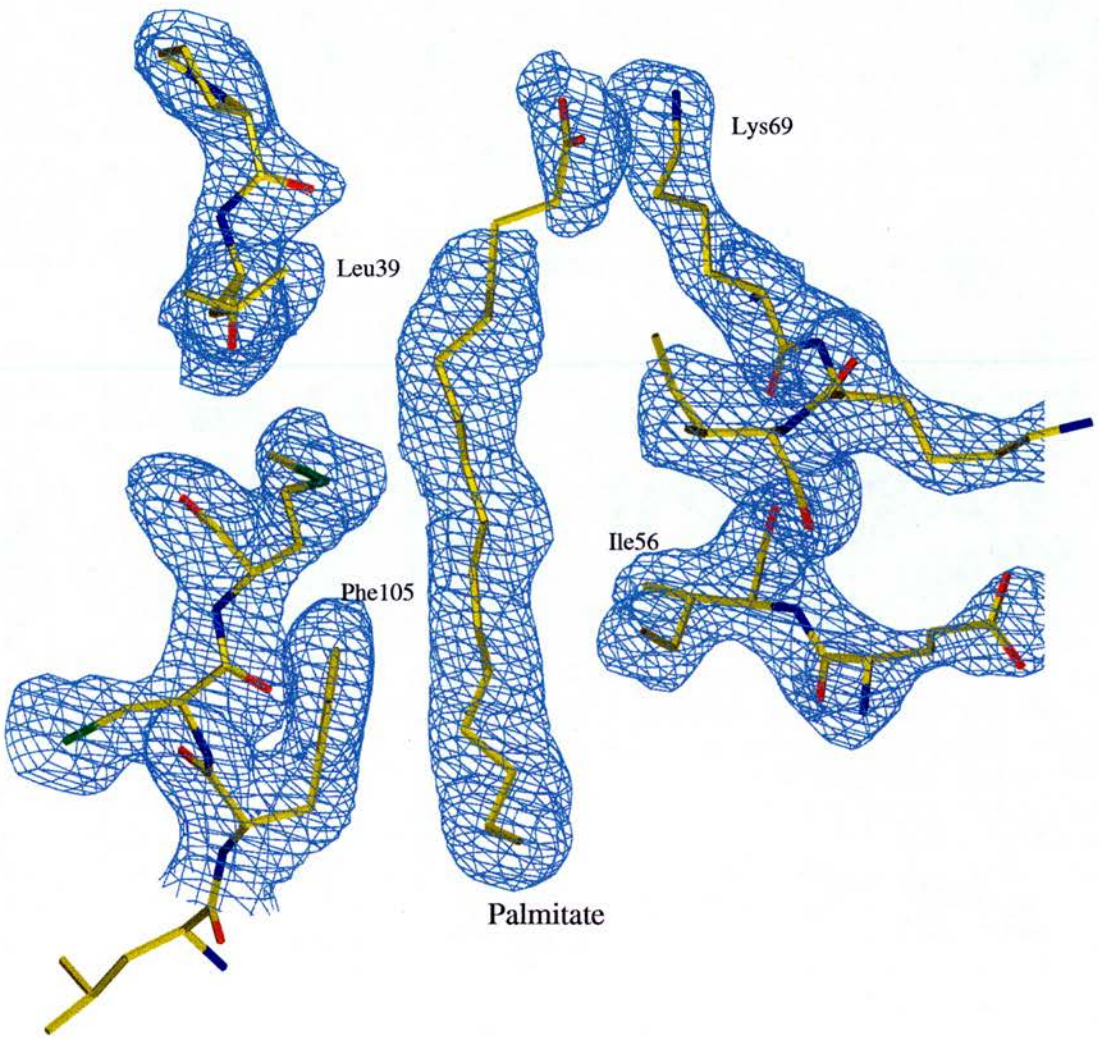


Figure 4–6.The (2Fo–Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 1.3 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) after refinement in SHELX and displayed in O (Jones et al., 1991). Compared to Figure 4–2, this density map shows the improvement in the density of palmitate after the refinement in SHELX.

The model of palmitate was taken from the well-refined structure of palmitate binding to fatty acid binding protein (PDB ID: 1MDC) and built into the density with O. The tail of palmitate was built towards the bottom of the calyx, the head being located near the mouth of the calyx.14 out of 16 carbon atoms were able to be fitted into the density with the exception of two carbon atoms near the head of palmitate (C2 and C3). After palmitate was modelled into the density, 10 cycles of positional and B-factor refinement were performed. A (2Fo-Fc) map was again calculated. The ligand density was very clear at this stage (See Figure 4-7).

All atoms were fitted into the map and the density of the -COOH group was well-defined. Following ligand building, the protein structure was completed by inserting three residues (Ile2-Thr4) into the N-terminal region, and two residues (His161-Ile162) into the C-terminal region. These residues were unclear in the lattice X structure and had been omitted.

The final R-factor is 20.4% and R_{free} is 24.0% for all reflections to 2.5Å resolution. The result of the final round of refinement result is summarised in Table 4-F.

Table 4-F. Summary of the refinement statistics of BLG complexed with palmitate structure

Summary of the refinement statistics	**Palmitate/BLG structure**
Resolution range (Å)	10-2.5
R_{final} (F>4σ)	20.4% (19.7%)
R_{free}[1] (F>4σ)	24.0% (23.7%)
Rms deviations from ideal geometry	
Bond lengths (Å)	0.004
Angle distance (Å)	0.018
mean B-factor (Å^2)	47.17

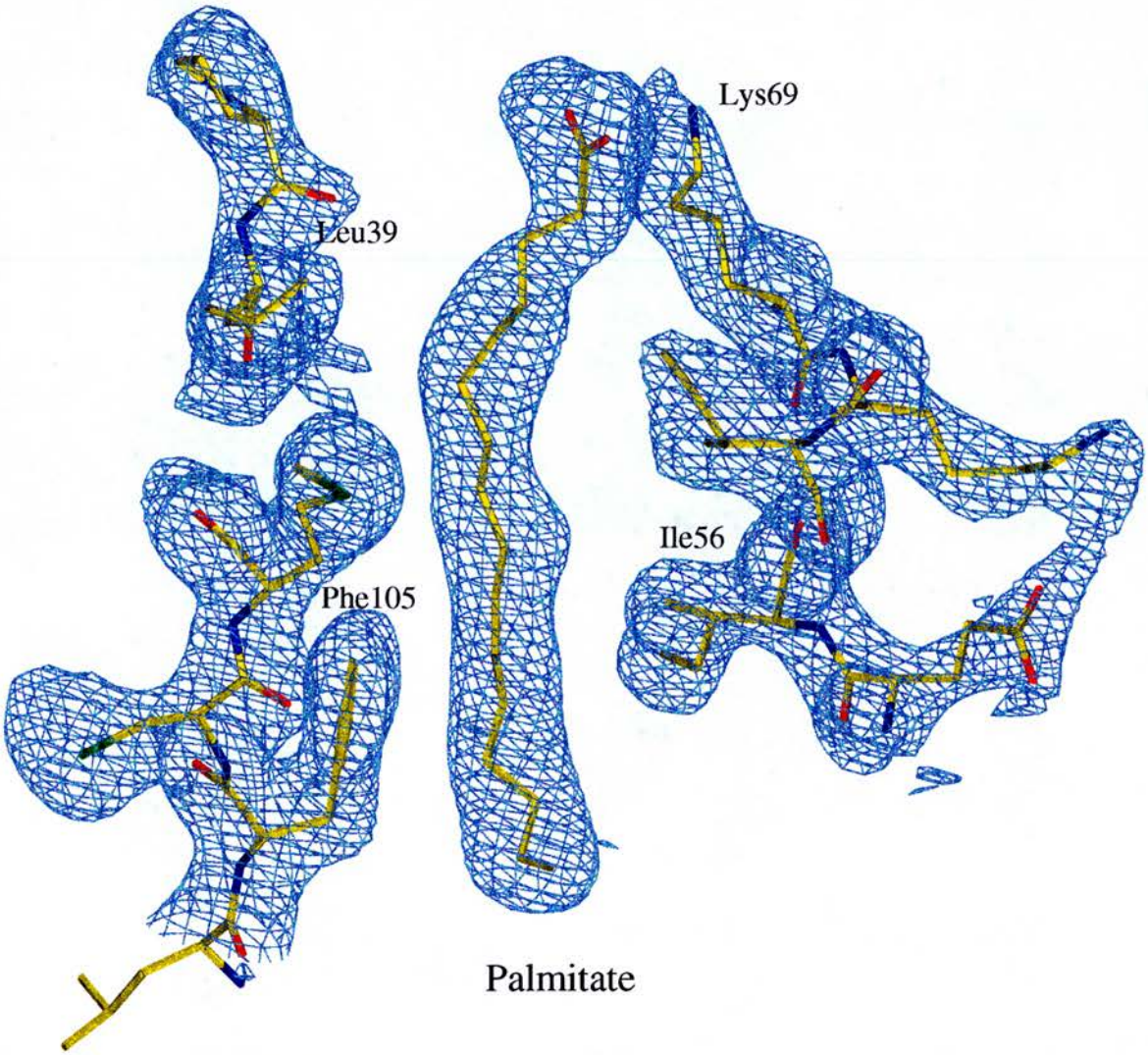


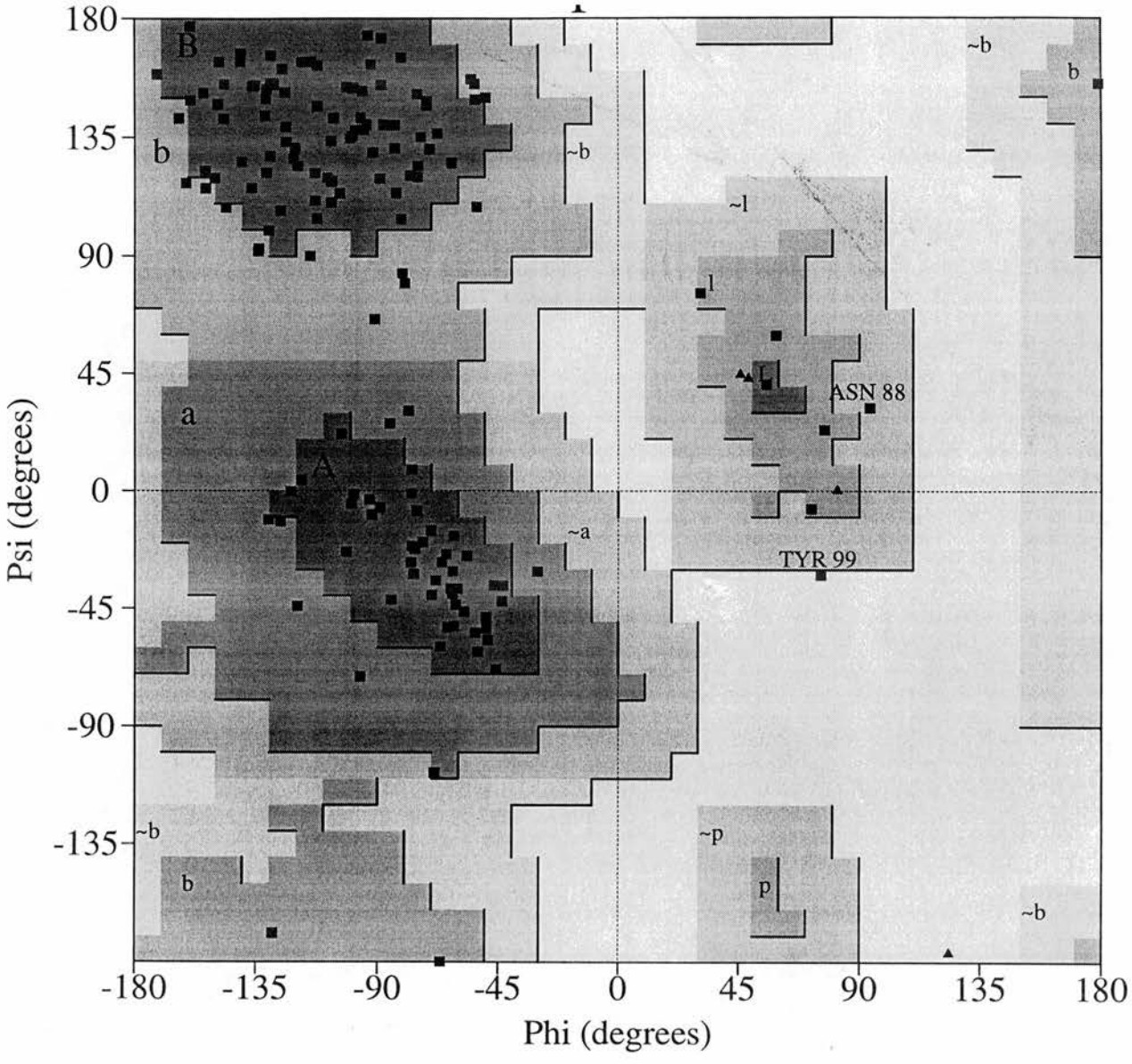
Figure 4-7.The (2Fo-Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 1.3 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) after building palmitate into the density and displayed in O (Jones et al., 1991). The density of palmitate is well defined.

The solvent content of the crystal is 50.78 %, with the Matthews number 2.52 $Å^3$/Da. There are 6 molecules in the unit cell, with 1 molecule per asymmetric unit. The electron density map of the 9 β-strands and an α-helix is very clear. However, N-terminal, C-terminal and some loops that connect β-strands have ill-defined density, especially the CD, EF and GH loops. Regions of disorder are commonly found in BLG structures (Brownlow *et al.*, 1997, Qin *et al.*, 1998a). These loops are around the open end of the β-barrel and are flexible, which resulted in difficulty in interpretation of the electron density map. However, it is as a direct result of the flexibility in these regions that the ligand is able to access the calyx without obstruction.

The structure has good geometry as reported by SHELX97 with rms deviation of bond length and angle distance, 0.004Å and 0.018 Å, respectively. The average B-factor for all protein atoms is quite high, 47.17 $Å^2$, but is in keeping with the values usually obtained for BLG.

The stereochemical quality of the structure was checked by CCP4 program PROCHECK (Laskowski *et al.*, 1993). The Ramachandran Plot is shown in Figure 4-8.

Essentially, all residues are in the allowed regions, expect Tyr99. Tyr99, which adopts a typical γ-turn conformation common to other lipocalin structures with the sequence of TDY, is well defined in the density map and is also in the disallowed region in the lattice X structure. One residue, Asn88, is in the generously allowed region. This residue is in the mobile EF loop which is a portal loop to reveal the interior of the calyx and the density map around this part is weak and poorly defined. PROCHECK was also used to assess agreement between the main-chain and side-chain parameters of the structure of palmitate bound to BLG with those of well-refined structures at a similar resolution (See Figure 4-9 and Figure 4-10). The structure of palmitate bound to BLG is represented by a solid square, while other reference structures are represented by the dark band. The central line is a least-squares fit to the mean trend as a function of resolution and the width of the band corresponds to a variation of one standard deviation about the mean.



Plot statistics

Residues in most favoured regions [A,B,L]	116	78.9%
Residues in additional allowed regions [a,b,l,p]	29	19.7%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.7%
Residues in disallowed regions	1	0.7%
Number of non-glycine and non-proline residues	147	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	8	
Total number of residues	161	

Figure 4-8. Ramachandran plot of the structure of BLG with bound palmitate, produced by PROCHECK (Laskowski *et al.*, 1993). Square symbols represent the non-glycine residues while triangle represent the glycine residues

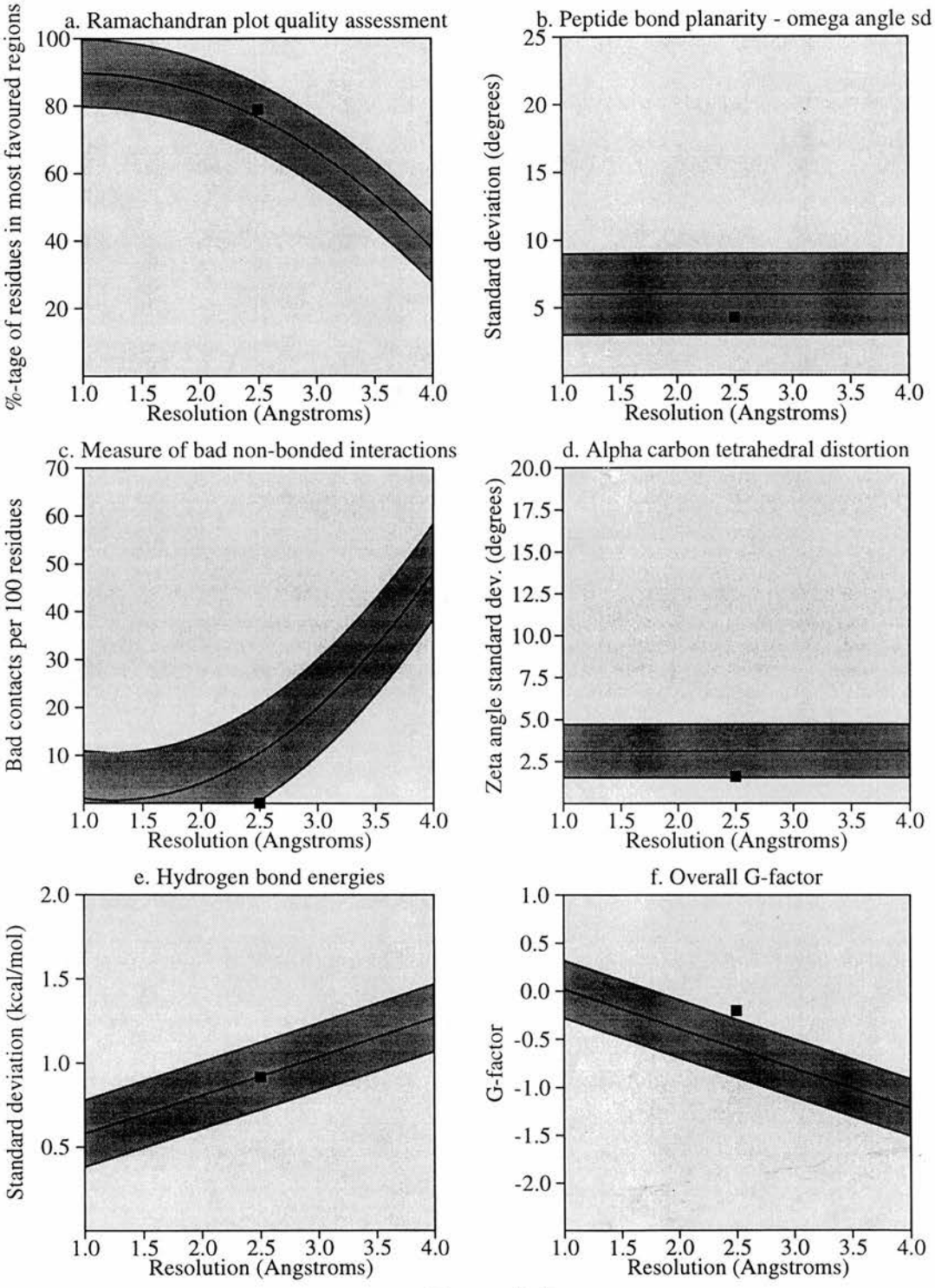


Figure 4-9. Plots of the comparison of the main chain parameters of the structure of BLG with bound palmitate with other structures solved to a similar resolution, produced by the program PROCHECK (Laskowski *et al.*, 1993).

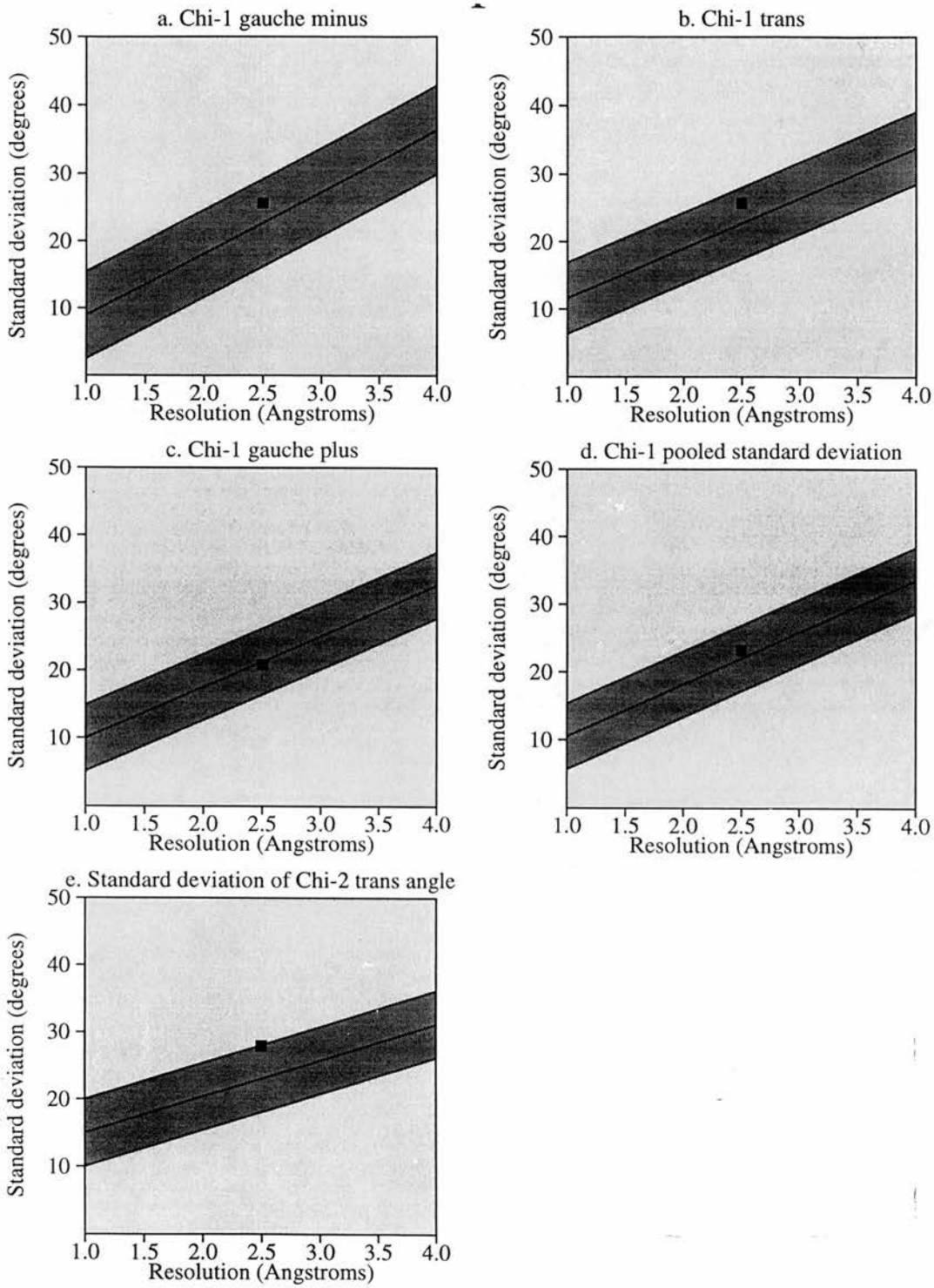


Figure 4-10. Plots of the comparison of the side chain parameters of the structure of BLG with bound palmitate with other structures solved to a similar resolution, produced by the program PROCHECK (Laskowski *et al.*, 1993).

The structure was of high quality since all main-chain and side-chain parameters are in the average region; two main-chain parameters, bad contact check and overall G-factor, are even better than the mean value.

4.2 Overall structure

The overall structure of palmitate bound to BLG is shown in Figure 4-11.
The structure of the host protein, BLG, is almost identical to the native BLG structure solved in the different lattice forms, triclinic (lattice X) and orthorhombic (lattice Y). The β-barrel of BLG is open to ligand binding. At the open end of the β-barrel, there are three very mobile loops, CD, EF and GH loops, and their flexibility allows the ligand to enter the binding site without obstruction. Inside the barrel is a hydrophobic cavity or calyx formed by the side chains of Leu10, Ile12, Val15, Val41, Val43, Leu46, Leu54, Ile56, Leu58, Ile71, Ala73, Ala80, Phe82, Ile84, Val92, Val94, Leu103, Phe105 and Met107. The palmitate binds inside the calyx and lies parallel to β-strands E with the hydrophobic tail stretching into the bottom of the calyx. At the open end of the barrel, the head of palmitate forms two hydrogen bonds with Lys60 and Lys69. The palmitate makes a total of 29 hydrophobic contacts with the 13 residues in the hydrophobic cavity over a distance of less than 4Å. The interatomic distances between palmitate and BLG are summarised in Table 4-G.
The distances between palmitate and BLG vary from 3.30 to 3.89Å. In particular, the side chain of Phe105 faces the tail of the palmitate and contributes most to the hydrophobic interaction (see Figure 4-12). From the statistics in Table 4-G, it is clear that the carbon atoms of the tail make more hydrophobic contacts with the protein than those of the head. Therefore, the carbon atoms at the tail with lower B-factor are stabilised by hydrophobic interactions while carbon atoms at the head with higher B-factors are more mobile due to weaker hydrophobic interactions with the protein. This phenomenon is also consistent with the density map of palmitate seen in Figure 4-2. The density of the tail is stronger and can be seen in the first (3Fo-2Fc) map (see Figure 4-2). The density of the head is rather weak and appears gradually in the later refinement stages.

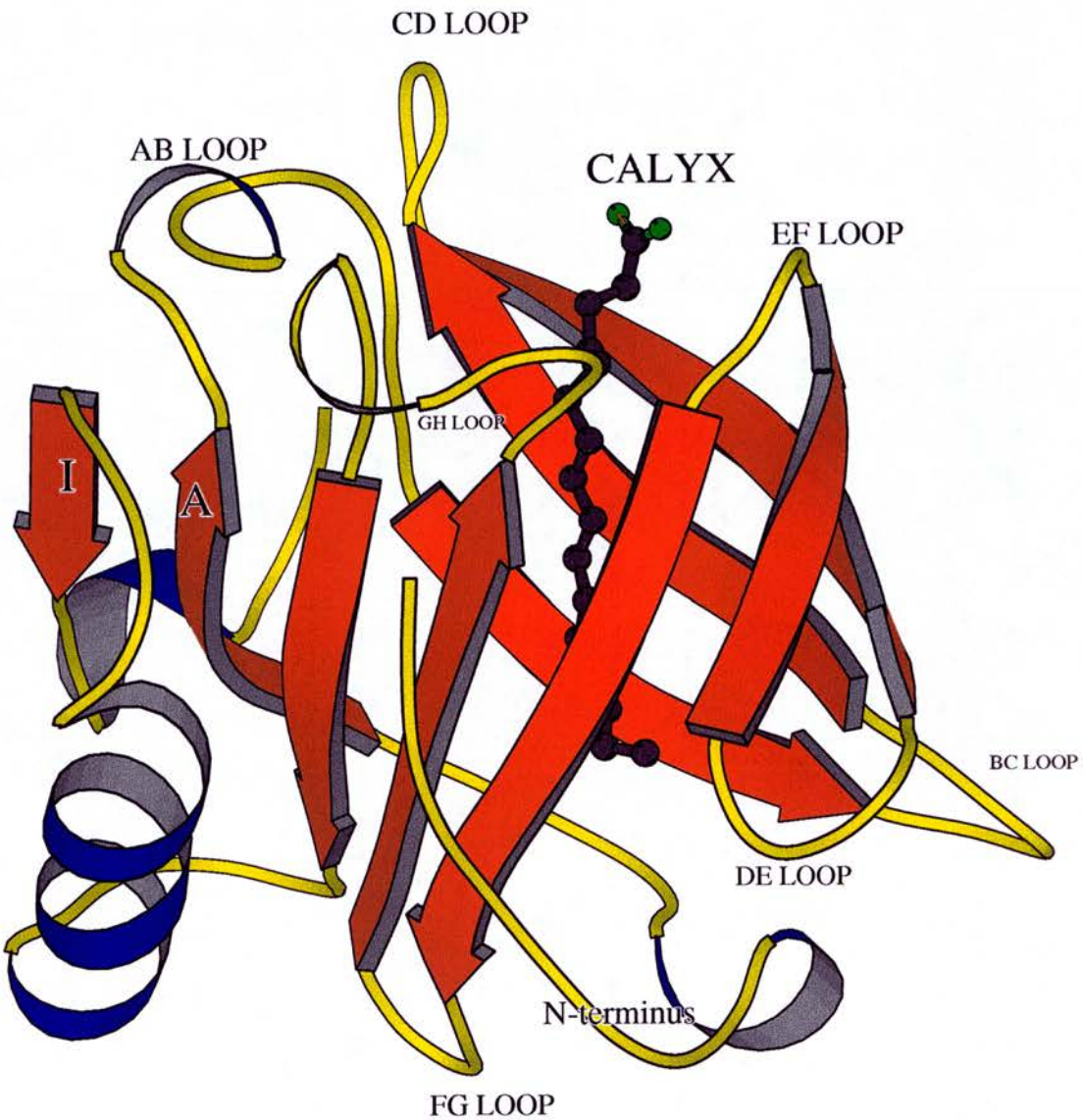


Figure 4-11.Ribbon diagram of the structure of palmitate bound to BLG. Palmitate binds inside the calyx.

Table 4-G. The interatomic distances between BLG and palmitate

	Palmitate atoms	BLG atoms	Distance (Å)
Hydrophobic interaction	C8	SD/Met107	3.56
	C7	SD/Met107	3.40
	C6	SD/Met107	3.89
	C13	CZ/Phe105	3.54
	C12	CZ/Phe105	3.57
	C11	CZ/Phe105	3.62
	C10	CZ/Phe105	3.81
	C13	CE1/Phe105	3.47
	C13	CE2/Phe105	3.53
	C12	CE2/Phe105	3.69
	C11	CE2/Phe105	3.30
	C10	CE2/Phe105	3.72
	C13	CD1/Phe105	3.41
	C13	CD2/Phe105	3.48
	C11	CD2/Phe105	3.89
	C13	CG/Phe105	3.44
	C15	CD2/Leu103	3.82
	C16	CG2/Val94	3.30
	C7	CD1/Ile84	3.50
	C5	CD1/Ile71	3.88
	C2	CE/Lys69	3.36
	C1	CE/Lys69	3.30
	C2	CD/Lys69	3.45
	C3	CD/Lys60	3.45
	C11	CD1/Ile56	3.83
	C16	CD2/Leu54	3.53
	C14	CD2/Leu46	3.69
	C14	CD1/Leu46	3.88
	C4	CG1/Val41	3.30
Hydrogen bond interaction	O1	NZ/Lys69	3.22
	O1	NZ/Lys60	2.72

There are also two strong hydrogen bonds involved in the ligand-protein binding. One oxygen atom of the carboxyl group of palmitate is hydrogen bonded to Lys60 and Lys69 with distances of 2.72Å and 3.22 Å, respectively. As mentioned in Qin's paper (Qin *et al.*, 1998b), Lys60 and Lys69 seem to be very important in the ligand binding affinity. The replacement of Lys69 with glutamate in pig BLG results in the failure of fatty acid binding to BLG (Frapin *et al.*, 1993). The replacement of Lys60 with glutamate in horse and donkey BLG also results in the failure of fatty acid binding (Perez *et al.*, 1993). It is thus tempting to speculate that any substitution of Lys60 and Lys69 will weaken the binding of fatty acid to BLG.

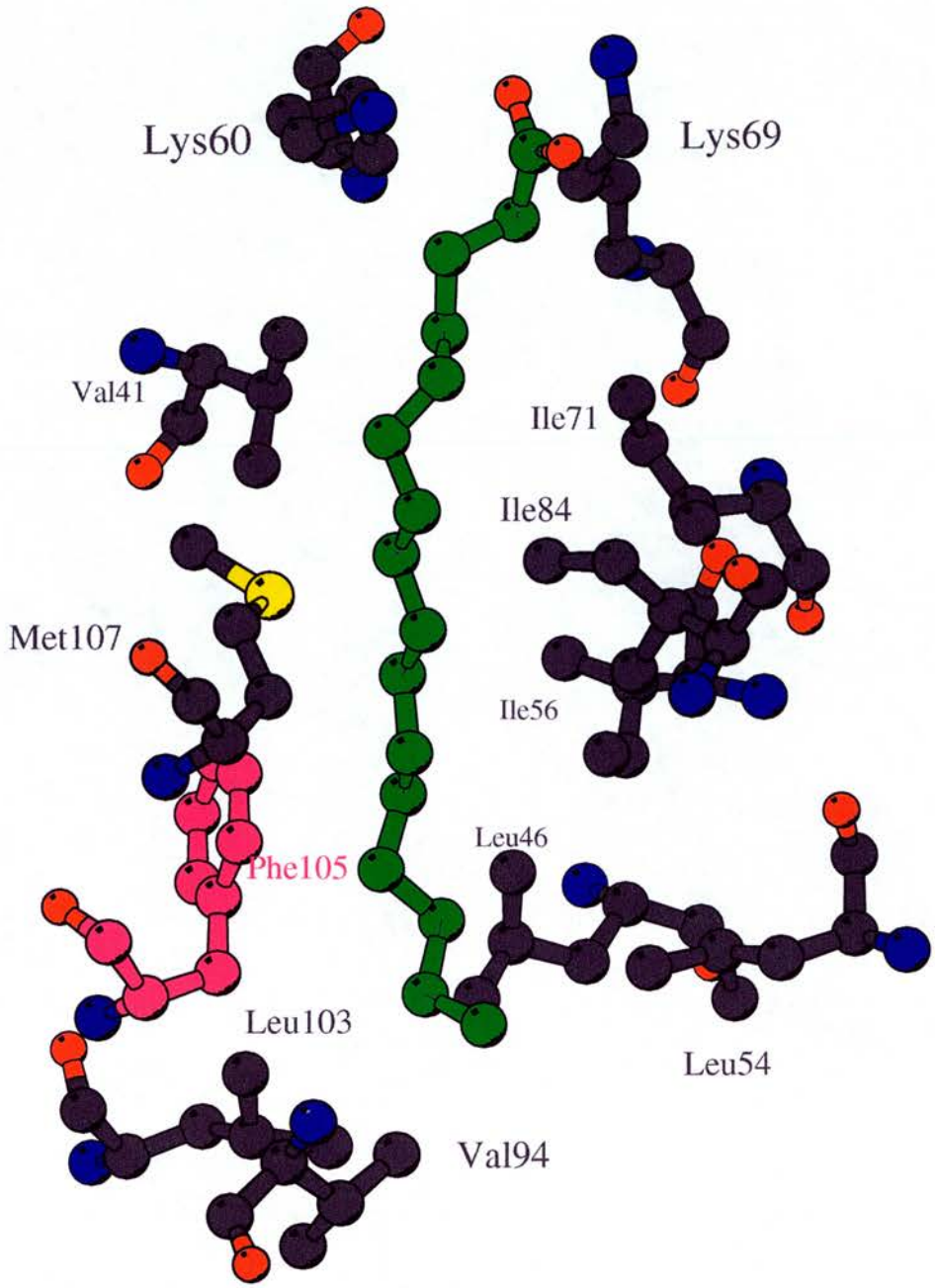


Figure 4-12.Diagram of the binding site of BLG with bound palmitate, produced by MOLSCRIPT (Kraulis, 1991). The residues formed the hydrophobic contacts with palmitate are labelled. The carbon atoms of palmitate are shown in green. The carbon atoms of Phe105 are shown in magenta.

4.3 The conformation of palmitate

In its bound conformation, palmitate bends slightly from a fully extended all-trans conformation in order to fit inside the calyx. There is a kink at C6. Careful comparison of the B-factors of the palmitate carbons within the calyx (C_8-C_{16}) shows they are similar in magnitude to those of the adjacent side chain atoms. Those nearer the opening of the calyx increase, presumably reflecting the larger volume of the cavity and hence the less specific nature of the binding in the region of the highly flexible EF and GH loops. The electron density for both these loops is not well-defined.

BLG contains two tryptophan, Trp19 and Trp61. Trp19, facing into the base of the calyx, is essentially inaccessible to solvent. Trp61 in the extremely mobile loop at the mouth of the hydrophobic calyx is exposed. As perturbation of the Trp fluorescence can be used to monitor binding of hydrophobic ligands to BLG, the distances from palmitate to the two Trp residues were determined directly in the program O. The closest approach of C2 is 10.36Å from Trp61 and C15 is 6.98Å from Trp19. It is possible therefore that there is perturbation of fluorescence signals from both Trp residues, although that at position 19 is in a significantly more restricted environment. Some papers have proposed Trp19 maybe be more important than Trp61 for the changes in both fluorescence and CD intensity by site-direct mutagenesis method (Cho *et al.*, 1994) and CD (Creamer, 1995). This is probably relative to the observation that Trp19 makes closer contact with ligand than Trp61.

4.4 The structure for the second data set

The structure of BLG co-crystallised with palmitate was used as the starting model to refine the structure from material produced by the alternative method as mentioned in section 2.3.2.3. Firstly, only the protein structure was used in refinement; the palmitate was omitted. Positional and B-factor refinement were performed using SHELX97. After 10 cycles of positional and B-factor refinement, a (2Fo-1Fc) map

was calculated using SHELXPRO and displayed with O. In the (2Fo-1Fc) map (see Figure 4-13), some density appeared in the expected binding site inside the calyx. The density of the tail of palmitate is visible, whereas that of the head is very weak. 100 H_2O were added using SHELXWAT and a (2Fo-1Fc) map was calculated again. The density map (see Figure 4-14) is clearer and the density around the head region of palmitate is stronger.

The R-factor also decreased from 28.0% to 23.6%. Palmitate was then built into density. After building, 20 cycles of positional and B-factor refinement were performed again and another 30 H_2O were added. The ligand density is very clear at this stage (see Figure 4-15).

All palmitate atoms were fitted into the map and the density of the -COOH group were well-defined. The structure was completed by manually building some residues of the flexible loops, for examples, Asn88, Glu89 and Pro113, into density.

The final R-factor is 23.3% and R_{free} is 28.5% for all reflections to 2.3Å resolution. The result of the final round of refinement result is summarised in Table 4-H.

Table 4-H. Summary of the refinement statistics from the second data set

Summary of the refinement statistics	**Second data set**
Resolution range (Å)	10-2.3
R_{final} (F>4σ)	23.3% (21.9%)
R_{free}[1] (F>4σ)	28.5% (26.7%)
Rms deviations from ideal geometry	
Bond lengths (Å)	0.006
Angle distance (Å)	0.02
mean B-factor (Å^2)	47.51

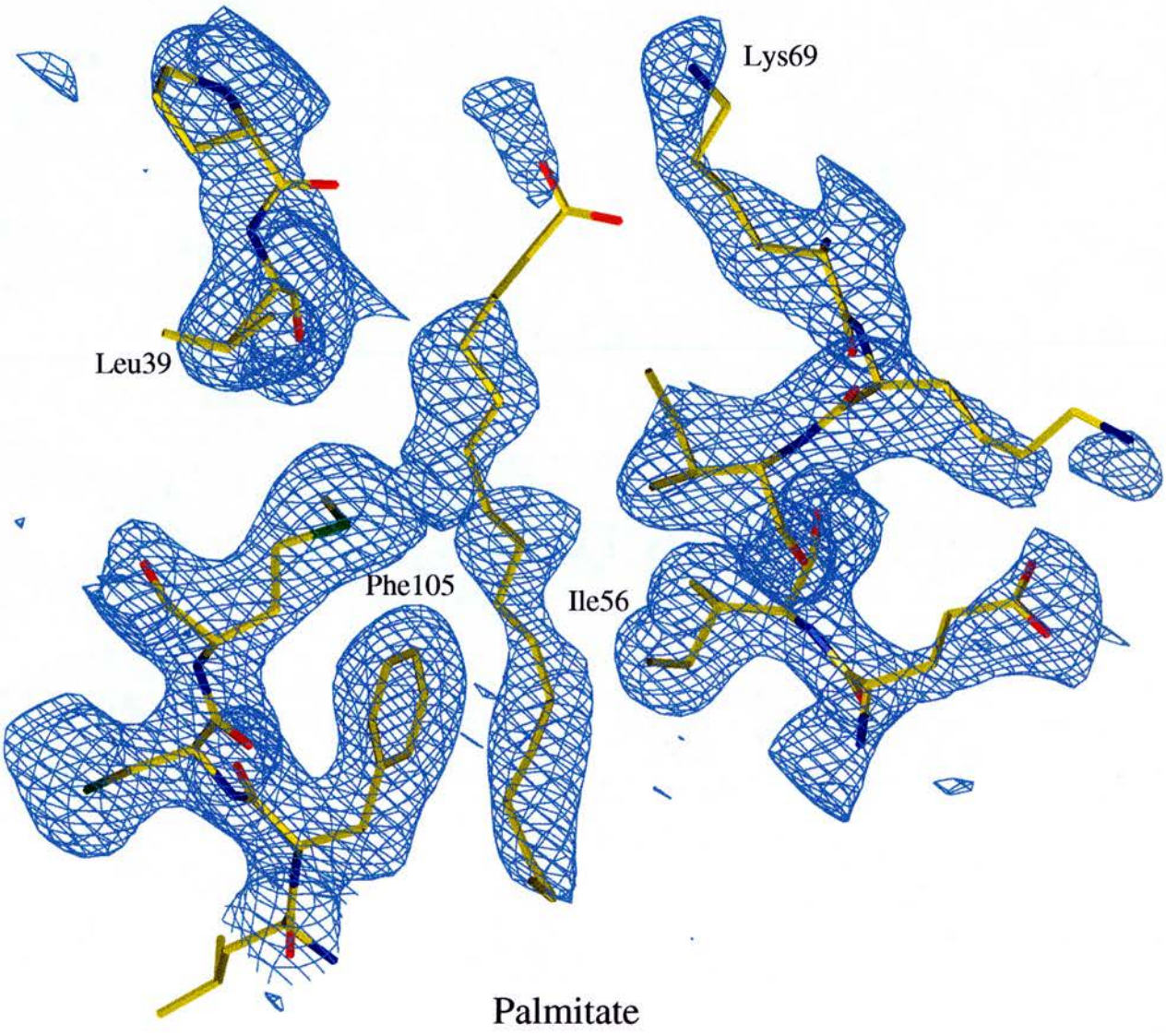


Figure 4-13.The (2Fo-Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 1.4 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) after positional and B-factor refinement in SHELX and displayed in O (Jones et al., 1991). Some density appeared in the expected binding site inside the calyx.

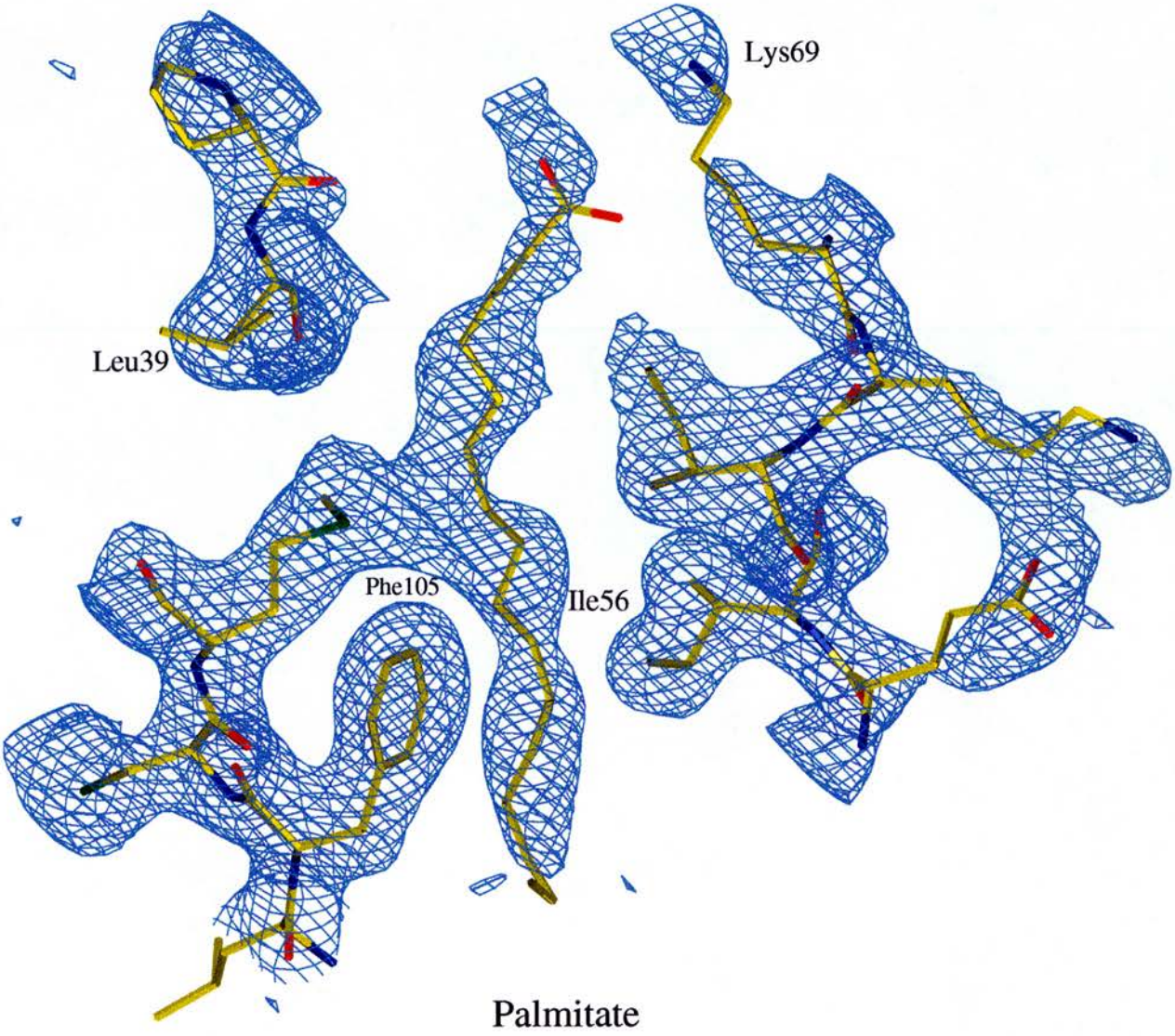


Figure 4-14.The (2Fo-Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 1.4 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) after the addition of water molecules in SHELX and displayed in O (Jones et al., 1991). Compared to Figure 4-13, this map shows the improvement in density of palmitate after the addition of water molecules.

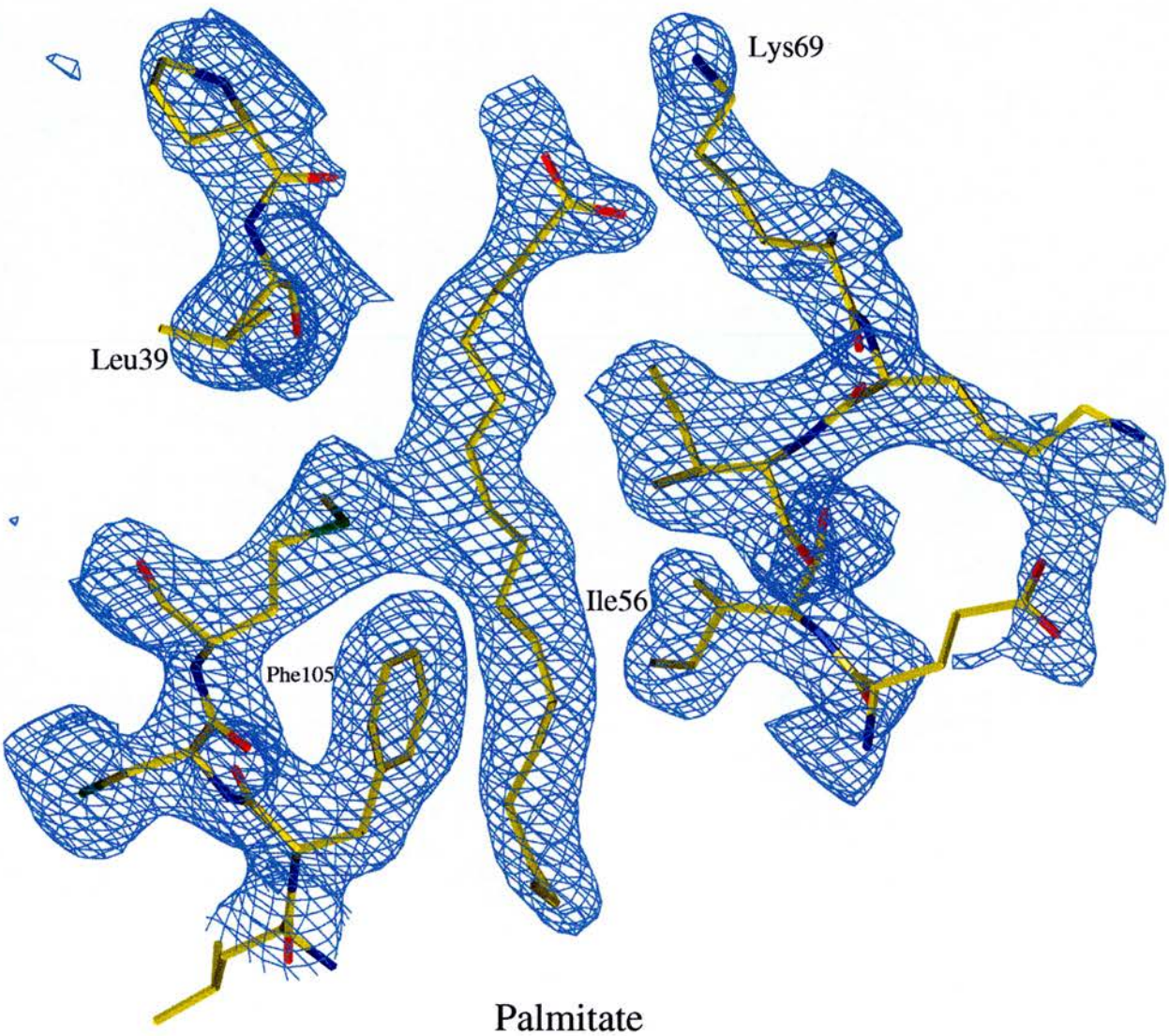


Figure 4-15.The (2Fo-Fc) electron density map around the bound palmitate with the refined structure superimposed. The map, contoured at 1.4 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) after building palmitate into the density and displayed in O (Jones et al., 1991). The density of palmitate is very clear in this map. Note the discrete density for Met107 and the palmitate.

The rms deviations of bond length and angle distance are 0.006Å and 0.02 Å, respectively. The average B-factor for all protein atoms is 47.51 $Å^2$. The stereochemical quality of the complexed structure assessed by PROCHECK is very similar to that of the structure from the first preparation. The geometry is again acceptable, with Tyr99 and Asn88 again being the exceptions in the Ramachandran plot.

It is clear that there is a bulge in the density associated with the position of Met107 (see Figure 4-15). The B-factors for the same palmitate C-atoms after refinement with the second data set were significantly larger than those of the adjacent side-chains but, when a single occupancy term for all ligand atoms was used in refinement, the B-value became similar to those of the surrounding side-chains. In the lattice Z crystal form, the two binding sites in the dimer are crystallographically identical. Analysis by gas chromatography as mentioned in section 2.2.2 showed that the complex had about 1 mol palmitate bound to 1 mol of dimeric protein. If the stoichiometry was one ligand per dimer, the expected occupancy would be 0.5 since on average, each site has half a palmitate molecule in it. The refined occupancy was 0.69. This is higher than 0.5, but because of the imprecise nature of the occupancy determination and the high correlation that exists between occupancy and B-factor, the result indicates a reasonable agreement with the solution study. In comparison, an identical calculation of the occupancy of palmitate in the first data set gave a value of 0.96. The bulge in the electron density at Met 107 is thus explained as the addition of two components for the Met 107 side chain – that of the native structure and that with bound ligand (see Figure 4-16).

What is not clear is why the different preparation protocols used should produce different palmitate to BLG dimer ratio. There is no obvious cooperativity between the monomers, whose binding sites, some 35Å apart, are approached from a direction away from the dimer interface.

One possible explanation is that first data set was collected from the crystal which was crystallised in the mother liquor super-saturated with palmitate as mentioned in section 2.3.2.2 while the second data set was from the crystal crystallised in the mother liquor containing no palmitate. Therefore, BLG was saturated with palmitate

once crystallised during the first preparation whereas some palmitate was dissociated from BLG to the mother liquor once crystallised during the second preparation. More palmitate thus bound to BLG in the first data set, which consequently results in the higher occupancy of palmitate compared to the second data set.

4.5 Comparison of the structure of Palmitate-bound BLG with other native BLG structures

Comparison of the structure of palmitate-bound BLG at pH 7.5 with the structure of the native lattice X at pH 6.5 reveals the major difference is in the movement of the EF loop (see Figure 4-16). In the lattice X structure, the EF loop is folded toward the calyx thus blocking the entrance to the calyx. In contrast, in the structure of palmitate bound to BLG, the EF loop is folded back to allow access for the binding of palmitate. It is the movement of the loop as the pH is raised from pH 6 to 7.5 that opens up the entrance to the calyx (see Figure 4-16).

The occlusion of, or at least hindrance to, the binding site at low pH may provide an explanation for the failure of the soaking experiments with the lattice X crystal form at pH 6.5. This result may be also relevant to the lower affinity binding of BLG to palmitate seen as the pH decrease from 8.7 to 6.5 (Spector and Fletcher, 1970). Spector measured the strength of palmitate binding to BLG over different pH range. This showed palmitate bound more tightly at pH 7.4 than pH 6.9. In addition, titration of BLG with palmitic acid, myristic acid and oleic acid under acidic conditions at pH 3.0 showed no binding of fatty acids as measured by the fluorescence enhancement assay (Frapin *et al.*, 1993). Therefore, it is tempting to conclude, at lower pH, the binding site is inaccessible to ligands because of hindrance by the EF loop.

The pH-dependant movement of the EF loop is in agreement with the study by Qin (Qin *et al.*, 1998a) as mentioned in section 1.1.7.

Apart from the EF loop, the side chain of Met107 near the kink at C6 also moves to allow the occupancy of palmitate (see Figure 4-16).

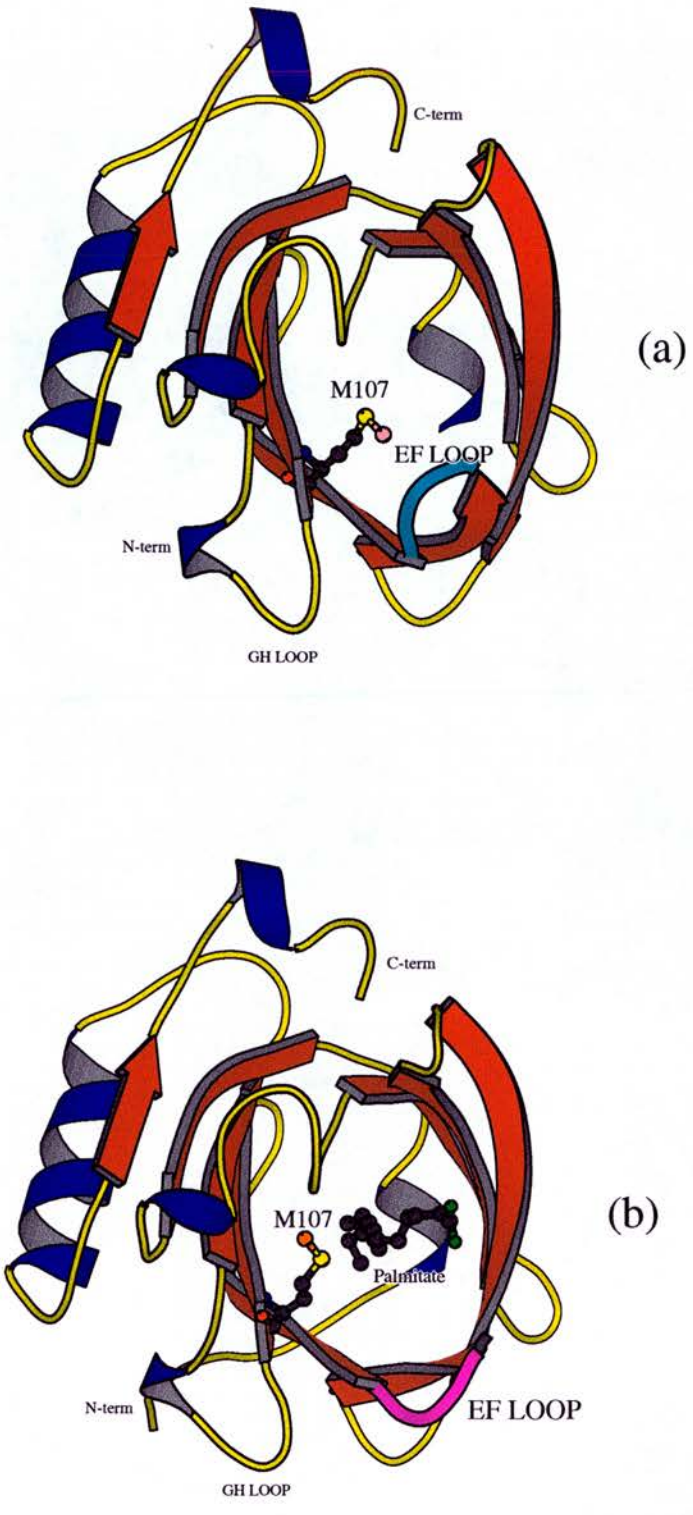


Figure 4-16.(a) Diagram of the native structure at pH6.5 (lattice X) showing the position of Met107 when palmitate is absent. (b) Diagram of the structure at pH7.5 (lattice Z) showing the movement of side chain of Met107 when palmitate binds. The movement of the EF loop can also be seen. There are no other significant movements between the bound and free forms of the protein.

Comparison of the structure of palmitate-bound BLG at pH 7.5 with the structure of the native lattice Y at pH 7.5 shows the binding sites of these two structures are identical (see Figure 4-17).

The exceptions are that the side chains of Lys60, Lys69, Ile71 and Met107 adopt different rotamers. As mentioned above, the side chain of Met107 moves away to allow the occupancy of fatty acid once the ligand enters the binding site. Lys60 and Lys69 are in the very mobile region. Their flexibility allows these two residues to move so that the side chains can make better hydrogen bond contacts when ligands, such as fatty acids, bind into the calyx.

Additionally, the EF loop in the lattice Y form also adopts the open conformation as seen in the complexed structure. It is therefore tempting to speculate that fatty acid will bind in the lattice Y form in the same manner as in the complexed structure in the lattice Z form.

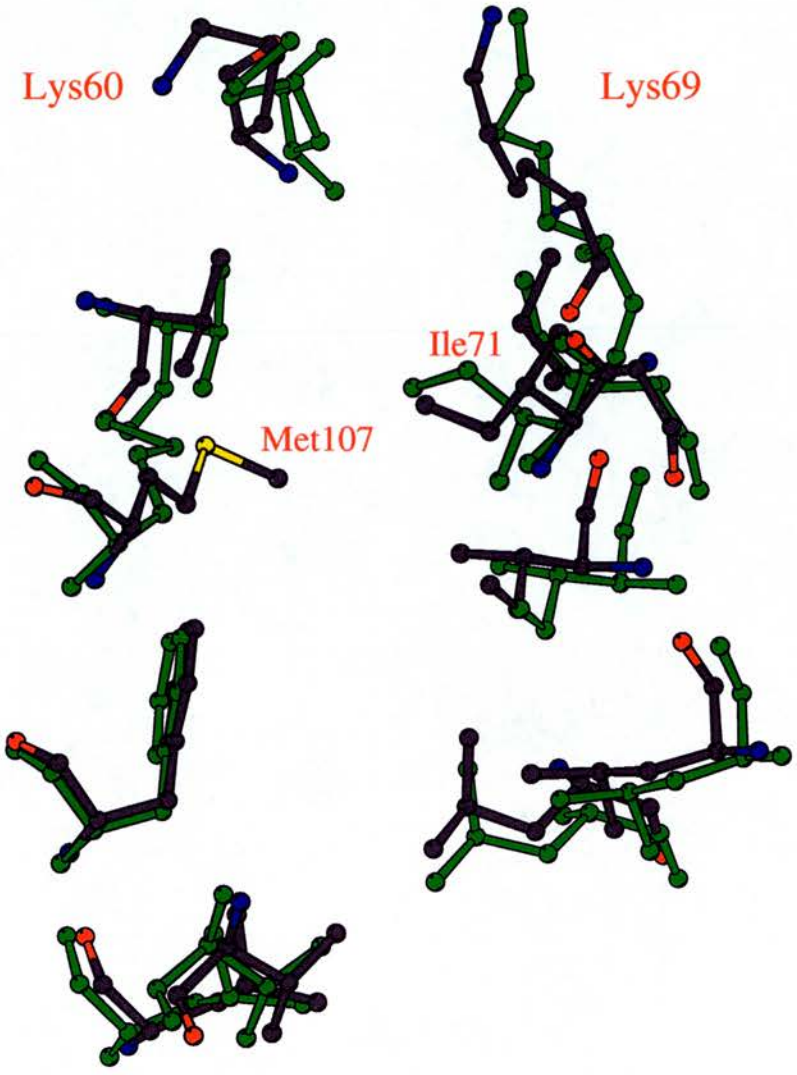


Figure 4-17.Diagram of the superimposition of the binding site of BLGY and that of palmitate bound to BLG structure (green). The labelled residues represent the residues that adopt the different conformations in these two structures.

4.6 Comparison of the structure of palmitate bound to BLG with that of 12-bromododecanoic acid bound to BLG

12-bromododecanoic acid binds to BLG in a very similar way to palmitate (Qin *et al.*, 1998b). The carboxyl group of 12-bromododecanoic acid is associated with Lys60 and Lys69 while the tail stretches into the interior of the calyx. Both adopt the almost fully extended conformation with a kink at C6 in palmitate and C3 in 12-bromododecanoic acid. By superimposing these two structures, it was found that the head of palmitate is one C-C bond length longer to the mouth of the calyx and three C-C bond lengths deeper into the base of the calyx. The palmitate makes a total of 29 hydrophobic contacts with the protein in a distance less than 4Å while 12-bromododecanoic acid makes a total of 12 hydrophobic contact with the protein over the same distance. The fewer hydrophobic contacts of 12-bromododecanoic acid with BLG provide the structural basis of the weaker binding strength of dodecanoic acid (lauric acid) than that of palmitate observed by fluorescence (Frapin *et al.*, 1993). The association constant of palmitate is 7 times higher than that of lauric acid. It also proved the assumption made by Perez (Perez *et al.*, 1995) that hydrophobic interactions are very important in the binding process. As for the structures of the BLG, these two are very similar, except in some mobile loops. The rms between corresponding backbone atoms of 145 residues in the structure of palmitate-bound BLG and that of 12-bromododecanoic acid-bound BLG is 0.403Å with the CD, EF and GH loops excluded. In the structure of palmitate bound to BLG, two loops at the mouth of the calyx, the EF loop and CD loop, slightly shift toward the ligand (see Figure 4-18)

The shift in the CD loop will allow Lys60 and Lys69 to make better hydrogen bonds with the carboxyl group of palmitate. The distances between O1 atom of palmitate and the side chains of Lys60 and Lys69 are 2.72Å and 3.22Å, respectively while the distances between O1 of 12-bromododecanoic acid and side chains of Lys60 and Lys69 are 4.5Å and 5.08Å, respectively. There are also other subtle differences between these two structures.

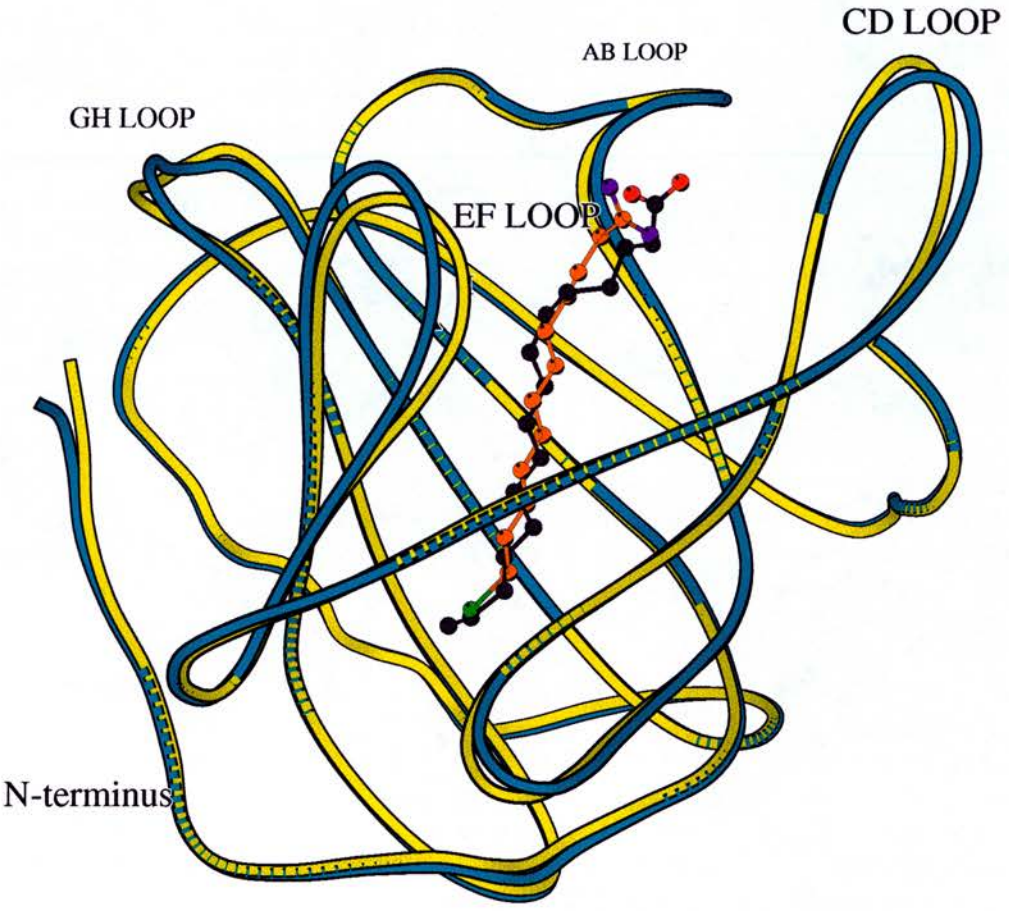


Figure 4-18.Diagram of the superimposition of palmitate bound to BLG and 12-bromododecanoic acid bound to BLG structures. The structure of BLG with bound palmitate is shown in yellow while the structure of BLG with bound 12-bromododecanoic acid is shown in cyan. The carbon atoms of palmitate are shown in black while the carbon atoms of 12-bromododecanoic acid are shown in orange.

The orientation of the carboxyl group of palmitate is different from that of 12-bromododecanoic acid. The orientation of the head of palmitate allows the carboxyl group to make better hydrogen bond contact with Lys60 and Lys69 along with the shift of the CD loop as discussed above. Moreover, the side chains of Met107, Ile84 and Ile71 adopt the different rotamers.

4.7 Comparison of the structure of palmitate bound to BLG with FABP structure

This gating of the inner binding site is in some way similar to that of the dynamic portal hypothesis in FABP where ligand access and binding are mediated by flexible regions of the protein backbone (Hodsdon and Cistola, 1997b). While in the structure of palmitate bound to BLG, discrete arrangements of the 'portal' loop EF exist and some nearby loops are very mobile. In the dynamic portal hypothesis, the disordered region creates a dynamic portal to allow entry of the fatty acid. Once the fatty acid binds to the protein, a variety of interactions induced by the ligand stabilise this flexible region and close the ligand entry portal. This fatty acid-induced ordering process may effectively lower the dissociation rate and increase the binding affinity. Any reaction that destabilises this disordered region will release the fatty acid. This hypothesis is relative to the observation by Qin (Qin *et al.*, 1998b). When 12-bromododecanoic acid binds to BLG, the EF loop is more rigid with smaller B-factors compared to the native structure. It confirms that ligand binding stabilises the flexible region in the entrance of the calyx. However, the phenomenon of the entry portal closing once ligand was bound to protein is not observed in BLG. In BLG, the EF loop adopts the open conformation with and without ligand binding.

The binding site in BLG is rather fully extended, but there is space for longer fatty acid molecules such as stearate and oleate to be accommodated within the calyx, with the carboxyl group making the same interactions with Lys 60 and Lys 69. The association constants for both acids are similar (Perez *et al.*, 1989, Frapin *et al.*, 1993). In the FABP family, palmitate is also observed in an extended form, although alternative conformations of bound fatty acids have been observed (Hodsdon and Cistola, 1997a, Sacchettini *et al.*, 1992). It has been found that if the carbon chain is

longer than 16 carbon atoms, the fatty acids, such as stearic acid or arachidonic acid, are likely to adopt a horse-shoe conformation whereas if the carbon chain is equal or shorter than 16 carbon atoms, the fatty acids always adopt the extended conformation.

4.8 The existence of the second binding site?

Little can be said about the existence of the second binding site so far although several lines of evidence point to the possibility of the existence of a second binding site. Retinol, the other well-known ligand that can bind to BLG, has been proposed to bind independently and simultaneously with fatty acid to BLG. For example, titration of native BLG, as well as retinol-bound BLG, by palmitic acid enhances the tryptophan fluorescence simultaneously and their association constants are quite similar, $Kd=1*10^{-7}$ M (n=0.93) and $Kd=2.6*10^{-7}$ M (n=0.9) for native BLG and retinol-bound BLG complex, respectively (Frapin *et al*, 1993). Also, using electron spin resonance spectroscopy, Narayan and Berliner (1997) have found 5-doxylstearic acid, a fatty acid analogue, binding at a site that is not perturbed by retinol and proposed the presence of at least two strong hydrophobic binding sites on BLG. Recently, Lange *et al* (1998) suggested the retinol and retinoic acid bind in the surface cleft of BLG instead of the interior calyx, using data from fluorescence resonance energy transfer (FRET) analysis. According to Lange's result, the experimental efficiency of fluorescence resonance energy transfer calculated from the two Trps (Trp19 and Trp61) and the ligands, retinol and retinoic acid, are in excellent agreement with the theoretical value calculated from the distances between each Trp and Lys141 as a reference point for the surface cleft. In contrast, the theoretical value calculated from the distances between each Trp and Lys70 as a reference point for the interior calyx is twice as large as the experimental value. It is needed to note that Lys70 in the model (Papiz *et al.*, 1986) used by Lange should be corrected to Lys69 in the revised model (Cabral, 1993; Brownlow *et al.*, 1997; Qin *et al.*, 1998a). The result obtained by Lange is consistent with the crystallographic analysis of the structure of retinol bound to BLG by Monaco (Monaco *et al.*, 1987), which suggested that retinol binds in the external groove. However, the proposed surface binding site

of retinol is in contradiction to the report of Cho *et al* (1994), which found that the residues of the central calyx contribute to the binding interaction as demonstrated by fluorescence study together with site-directed mutagenesis. Thus, higher resolution data from the retinol/BLG complex is required to facilitate further investigation into the precise nature of the retinol binding site. This would also allow clearer identification of second binding site of BLG.

5. Structure determination of CYP3

This chapter will describe the crystallographic work on the native structure of CYP3 and the structure of Ala-Pro bound to CYP3. Section 5.1 will contain the refinement of the CYP3 structure, the examination of the detail of the structure and the comparison with the CYPA structure. Section 5.2 will contain the refinement of the structure of Ala-Pro bound to CYP3, the analysis of the binding site and the calculation of the dissociation constant *"in crystallo"* by x-ray crystallography.

5.1 Native Cyp3 structure

5.1.1 Refinement

The initial structure was solved by Dr. Paul Taylor in Prof. Walkinshaw's group in Edinburgh (Taylor, 1998a) using molecular replacement with the refined human CYPA structure (space group $P2_12_12_1$: a=36.4Å, b=60.7Å, c=72.2Å, $\alpha=\beta=\gamma=90°$, PDB ID=1CWA) as the search model. Data used were to a resolution of 1.9Å and collected at the synchrotron. These data were from a crystal grown in the presence of 3.5%DMSO, which has been shown to bind to cyclophilin by x-ray crystallography (Kontopidis, 1999). Native data to 1.8Å was collected on the home source. This data set was used for further refinement. The native structure produced from this data set is discussed in this thesis.

The same refinement procedure mentioned in the section 2.7.3.2 was followed. Positional and B-factor refinement were performed using SHELX97. After 15 cycles of positional and B-factor refinement, H_2O molecules were added using SHELXWAT. 110 H_2O molecules were added and the R-factor decreased from 30.62% to 23.67% while R_{free} decreased from 36.07% to 29.45%. A (2Fo-Fc) map was calculated using SHELXPRO and displayed with program O to ensure the model is well-fitted in the density map. The density map was very clear and most of the residues were well defined in the map, except some residues in the N-terminal region, C-terminal region and the mobile loop, which will be discussed later. It was shown in the map that the side chain of Val140 adopted the different conformation

from the original model. Therefore, the side chain of Val140 was moved to fit into the density. Following the reposition of the side chain of Val140, several cycles of positional and B-factor refinement were performed again and more H_2O molecules were also added. The R and R_{free} of each refinement stage are summarised in Table 5-A.

The final R-factor is 20.81% and R_{free} is 28.47% to 1.8Å. The results from the final cycle of refinement are summarised in Table 5-B.

Table 5-A. The refinement process of CYP3

Refinement step	Resolution range(Å)	R (%)	R_{free} (%)
Positional and B-factor refinement	10-1.8	30.62	36.07
Adding 111 H_2O	10-1.8	23.67	29.45
Positional and B-factor refinement after moving side chain of Val140 into density	10-1.8	23.27	29.17
Adding another 150 H_2O	10-1.8	20.81	28.47

Table 5-B. Summary of the refinement statistics of CYP3 structure

Summary of the refinement statistics	**Native**
Resolution range (Å)	10-1.8
R_{final} (F>4σ)	20.81% (19.69%)
R_{free}[1] (F>4σ)	28.47% (26.61%)
Rms deviations from ideal geometry	
Bond lengths (Å)	0.031
Angle distance (Å)	0.026
mean B-factor (Å^2)	30.37

The solvent content of the crystal is 59.32 %, with a Matthews coefficient of 3.05Å^3/Da. The space group is $P4_12_12$ and the unit cell is a=b=60.60Å and c=123.1Å. There are 8 molecules in the unit cell, with 1 molecule per asymmetric unit. Most of the residues are well defined in the electron density map, the exceptions being 2 residues in the N-terminal region (Met1-Ser2), 2 residue in the C-terminal region (Lys172-Ala173) and 2 residues in the loop (Gly87-Glu88). The B-factors of

these disordered residues are very high and the electron density around these regions is weak and discontinuous.

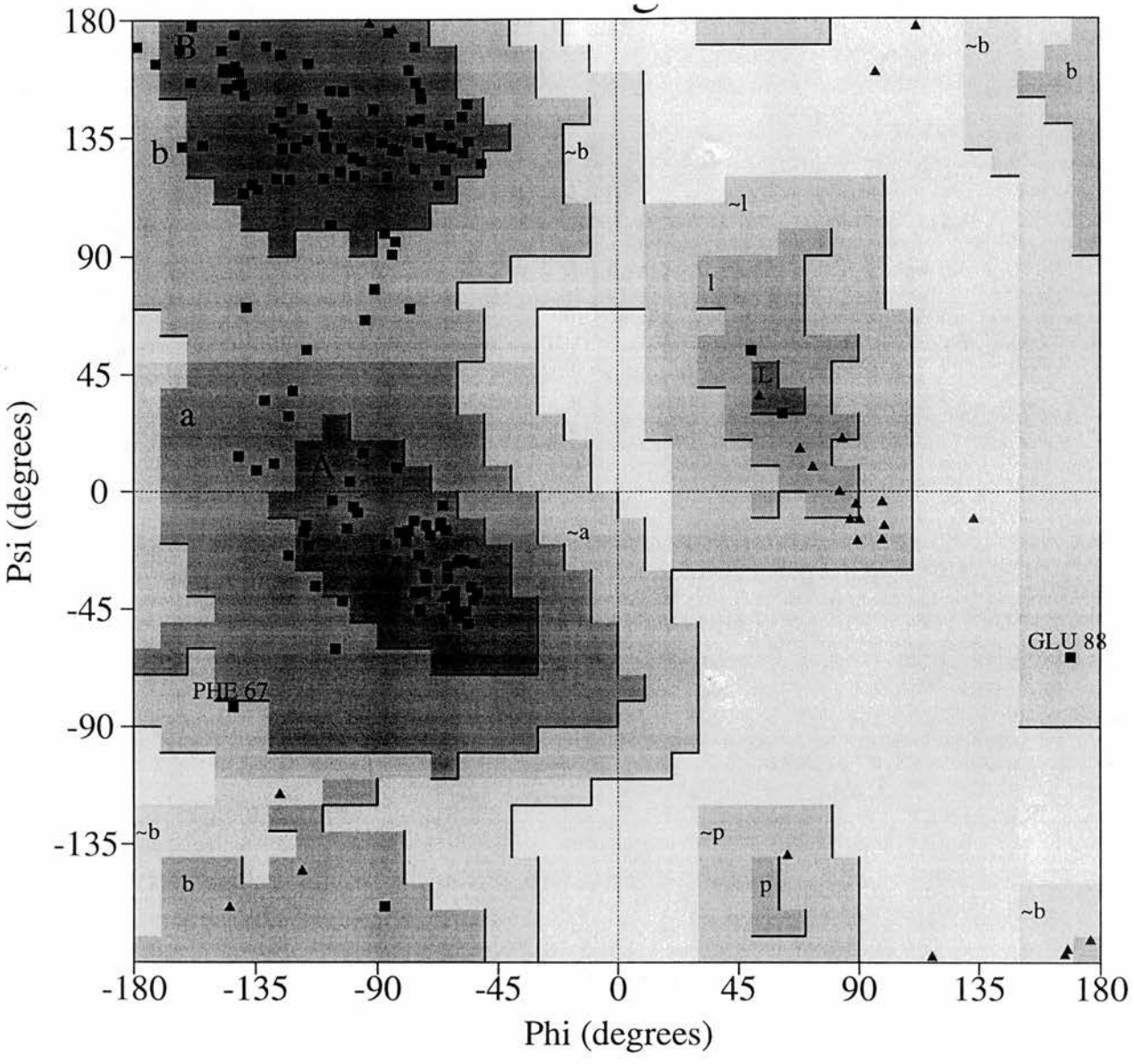
The structure has good geometry as reported by SHELX97 with rms deviations of bond length and angle distance, 0.031Å and 0.026 Å, respectively. The average B-factor for all protein atoms is 30.37 $Å^2$.

The stereochemical quality of the structure was checked by the CCP4 program PROCHECK (Laskowski *et al.*, 1993). The Ramachandran Plot is shown in Figure 5-1.

No residues are in the disallowed regions. Two residues, Phe67 and Glu88, are in the generously allowed regions. After carefully checking the density map, Phe67 is well defined in the density map as shown in Figure 5-2 while Glu88 is disordered and the density map is too weak to build this residue accurately.

PROCHECK was also used to assess the agreement between the main-chain and side-chain parameters of the CYP3 structure with those of well-refined structures at a similar resolution (See Figure 5-3 and Figure 5-4). In Figure 5-3 and Figure 5-4, the CYP3 structure is represented by a solid square, while other reference structures are represented by the dark band. The central line is a least-squares fit to the mean value as a function of resolution and the width of the band corresponds to a variation of one standard deviation about the mean.

From these statistics, it is clear the structure is of acceptable quality as the stereochemistry is in good agreement with what is expected for a structure at this resolution.



Plot statistics

Residues in most favoured regions [A,B,L]	112	81.8%
Residues in additional allowed regions [a,b,l,p]	23	16.8%
Residues in generously allowed regions [~a,~b,~l,~p]	2	1.5%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	137	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	7	
Total number of residues	172	

Figure 5-1. Ramachandran plot of the structure of CYP3, produced by PROCHECK (Laskowski *et al.*, 1993). Sqaure symbols represent the non-glycine residues while triangles represent the glycine residues

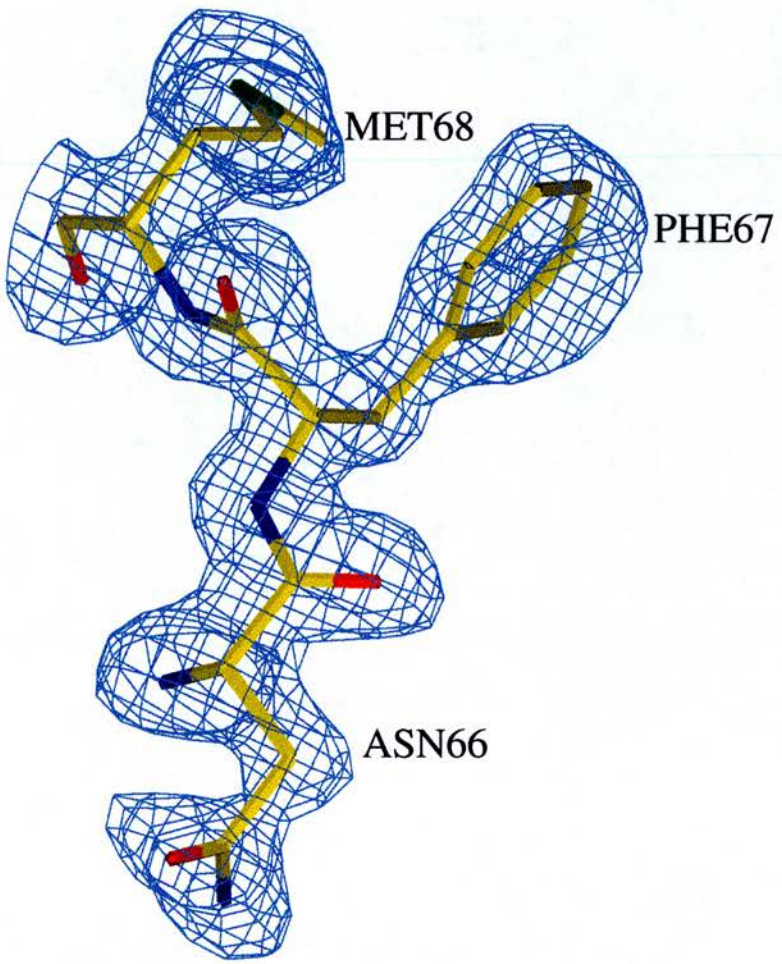


Figure 5-2.The (2Fo-Fc) electron density map around Phe67 with the refined structure superimposed. The map, contoured at 1.6 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) and display in O (Jones et al.,1991) Phe67 is well defined in the density map although it is in a disallowed region.

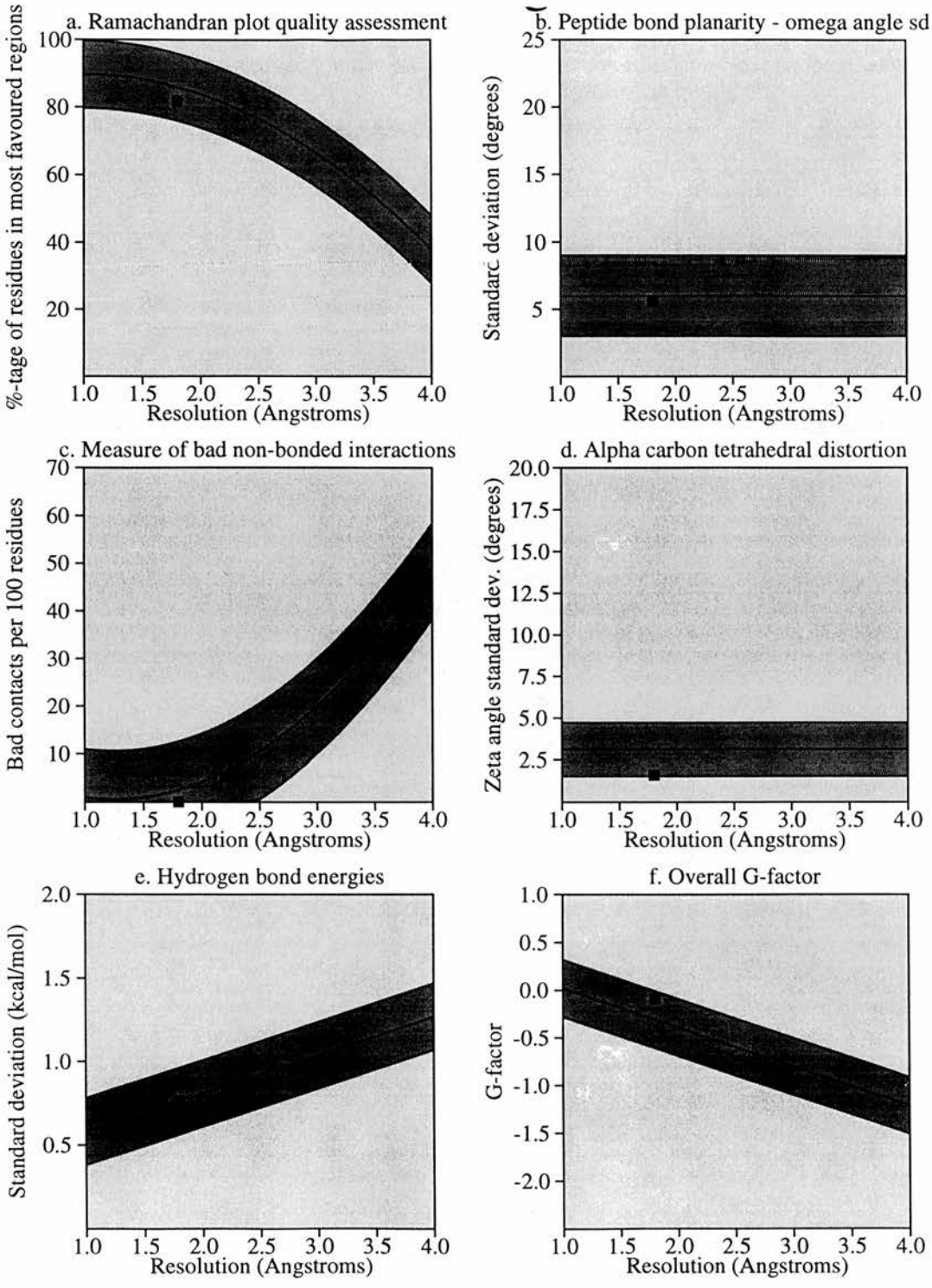


Figure 5-3. Plots of the comparison of the main chain parameters of CYP3 structure with other structures solved to a similar resolution, produced by the program PROCHECK (Laskowski *et al.*, 1993)

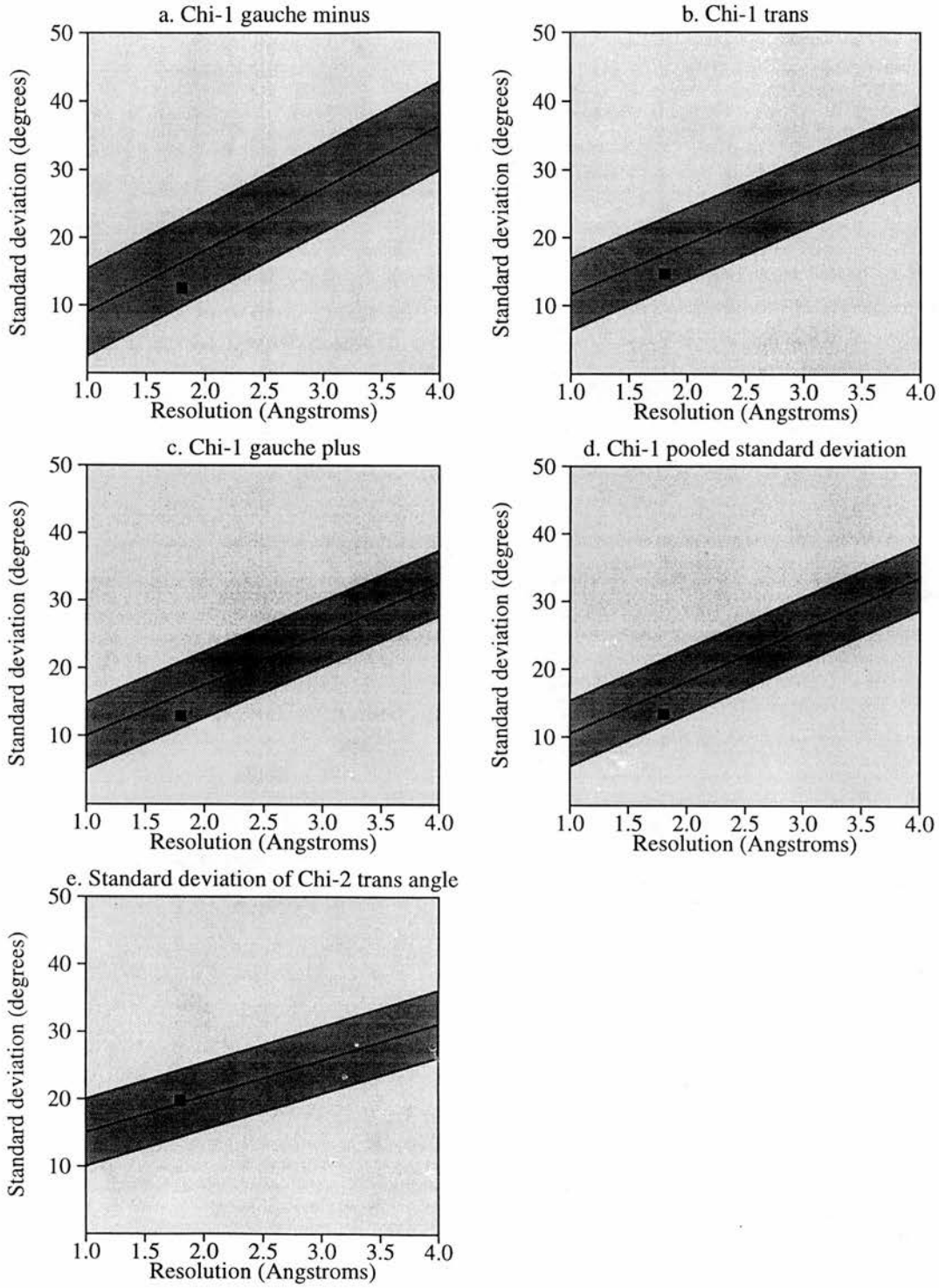


Figure 5-4. Plots of the comparison of the side chain parameters of CYP3 structure with other structures solved to a similar resolution, produced by the program PROCHECK (Laskowski *et al.*, 1993)

5.1.2 Overall structure

The overall structure of CYP3 is very similar to that of hCYPA with the exception of the inserted loop. The structural comparison of CYP3 with CYPA will be discussed in section 5.1.3.

In CYP3, there are 4 free cysteine residues. Cys122 and Cys163 are too far apart to form a disulphide bridge. However, Cys40 and Cys168 are very close, with 5.38 Å between 2 sulphur atoms, but fail to form the expected disulphide bond. Torsion angles Chi1 (N, CA, CB, S) of the 2 residues are -72° for Cys40 and 53° for Cys168. Modelling has shown that a rotation in the Chi1 torsion angle of both residues (-134° for Cys40 and -167° for Cys168), would enable the formation the disulphide bridge. This rotation changes the original staggered conformation to an alternatively favoured staggered conformation and leads to an S-S distance of 2.1Å with reasonable contacts with the surrounding residues.

From the clear electron density map around Cys40 and Cys168 (see Figure 5-5), there is no doubt CYP3 contains reduced cysteine residues. The reason why these two cysteines do not form a disulphide bridge is still unknown although this phenomenon is also seen in the cyclophilin from the parasitic nematode *Brugia malayi* (Taylor *et al.*, 1998b). It is probably that these unusual reduce forms of cysteine residues may provide a signalling mechanism in response to oxidative stress. In an oxidising environment, the formation of a disulphide bridge may induce a conformational change and this conformational change may provide a dynamic signal to activate other enzymes in the cell. Moreover, there is a metal ion coordinate site formed by the sulphydryl groups of the reduced Cys40 and Cys168 together with the imidazole side chain of His54. Interestingly, His54 is conserved in all divergent loop cyclophilin and probably plays an important role in this putative redox signalling mechanism.

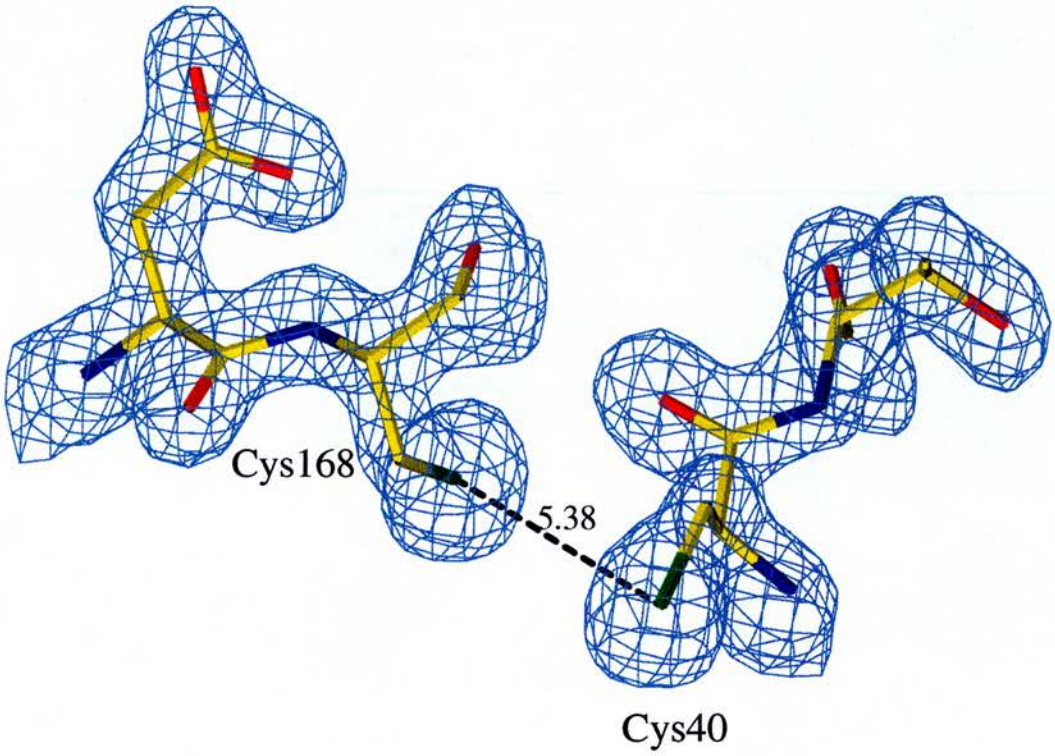


Figure 5-5.The (2Fo-Fc) electron density map around Cys40 and Cys168 with the refined structure superimposed. The map, contoured at 1.8 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) and displayed in O (Jones et al., 1991). The clear density shows that both cysteines are present in the reduced form.

5.1.3 Comparison of CYP3 with CYPA

CYP3 and CYPA have 67% sequence identity. The overall structures of both are very similar (See Figure 5-6).

The rms deviation between the corresponding backbone atoms of the 162 residues in CYPA and CYP3 is 0.469 Å; the 7 residues in the additional loop and 3 residues in N-terminal are excluded. The most distinctive structural feature between these two proteins is the additional loop found in CYP3 (see Figure 5-7).

Sequence alignment shows this additional loop containing 7 inserted residues (48KSGKPLH54), starting from Lys48 in CYP3 sequence. Comparison of these two structures shows divergence actually begins at Gly45. The loop is held by a tight and specific network of hydrogen bonds from the side chain of Glu83 (see Figure 5-8). The carboxylate group of Glu83 forms three strong hydrogen bonds with Lys48 and Ser49. The hydrogen bond distances are 2.79Å, 3.10Å, 3,57Å between N/Lys48 and OE2/Glu83, N/Ser49 and OE1/Glu83, OG/Ser49 and OE1/Glu83, respectively.

It is proposed that the role of Glu83 is to lock the loop into a particular β-turn conformation. This is a very important feature of the protein structure since such motifs are usually involved in protein-protein recognition, receptor binding and postranslational modification (Rose *et al.*, 1985). 4 residues (47GKSG50) of the loop form the β-turn. Lys48 and Ser49 are hydrogen bonded to Glu83 as mentioned above. In β-turns, the oxygen of the carbonyl group of the first residues is hydrogen bonded to the nitrogen in NH group of the fourth residue. As shown in Figure 5-9, the oxygen of Gly47 is hydrogen bonded to the nitrogen of Gly50 with a bond distance 2.77Å. Moreover, because ϕ and ψ of Lys48 and Ser49 are (-58, -26) and (-71, -12), this β-turn is assigned to type I, the most common β-turn type (Wilmot and Thornton, 1988).

The binding site of CYP3 is almost identical to that of CYPA, except that the side chains of Arg62, Met68 and Lys125 are slightly shifted (See Figure 5-10).

It is tempting to speculate that all ligands which bind to CYPA will bind to CYP3 and that the isomerase activity of CYP3 should be very similar to that of CYPA. Actually, the isomerase activity of CYP3 inhibited by the immunosuppressive drug cyclosporin has been determined with an IC50 value of 16nM (Dornan, 1999b) while that of hCYPA is 19nM. (Liu *et al*, 1990).

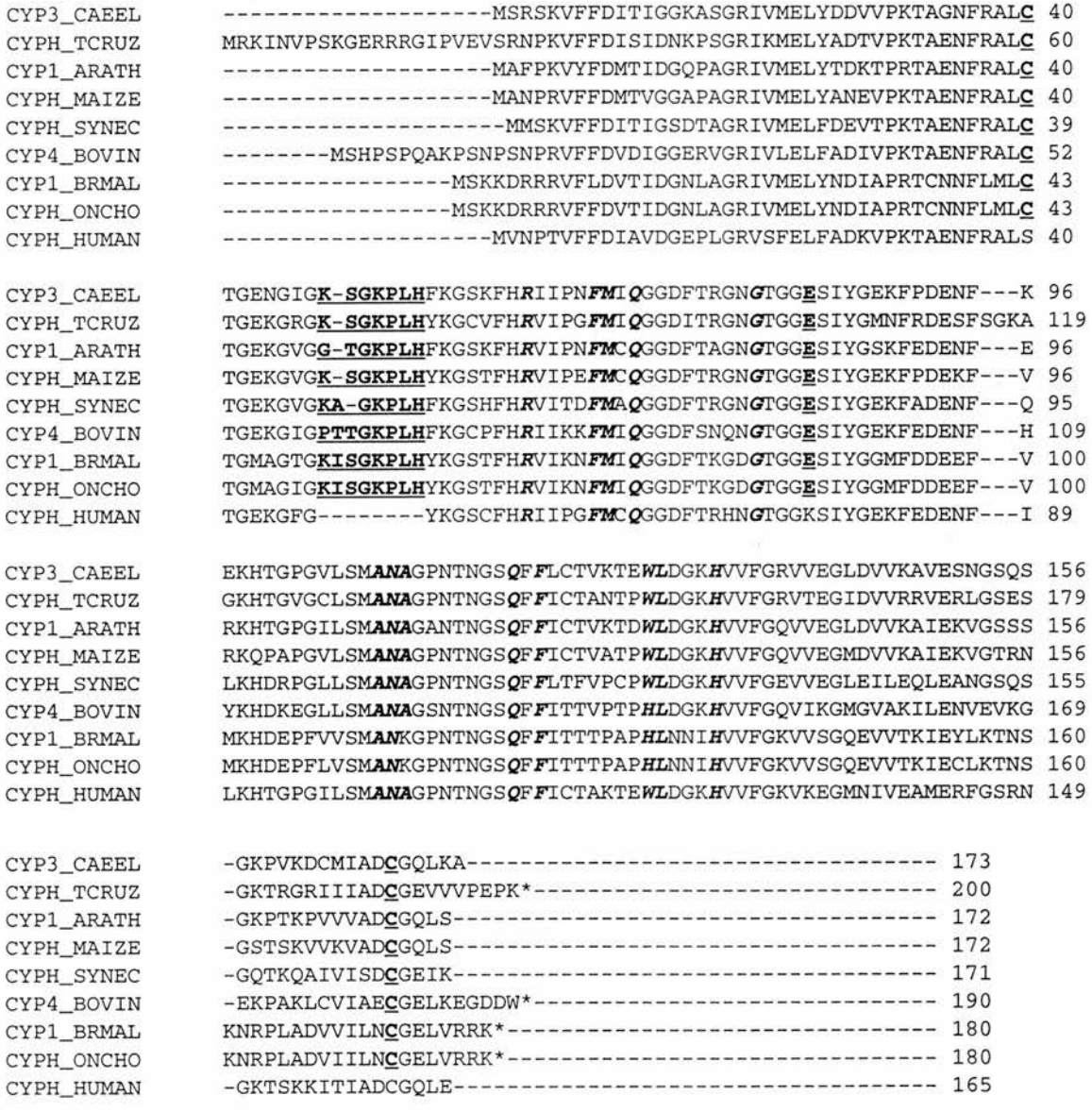


Figure 5-6. Sequence alignment of the cyclophilins, produced by CLUSTALW (Thompson *et al.*, 1994). Abbreviation: CYP3_CAEEL for CYP3 from *C. elegans*; CYPH_TCRUZ for cyclophilin from *T. cruzi*; CYP1_ARATH for CYP1 from *A. thaliana*; CYPH_MAIZE for cyclophilin from *Z. mays*; CYPH_SYNEC for cyclophilin from *Synechocystis sp*; CYP4_BOVIN for bovine CYP40; CYPH_BRMAL for cyclophilin from *B. malayi*; CYPH_ONCHO for CYP1 from *O. volvulus*; CYPH_HUMAN for human CYPA. Inserted residues forming the "additional loop" seen in all the aligned sequence with the exception of human CYPA are in bold and underlined. Cysteine, histidine and glutamate residues conserved within this subgroup of cyclophilin are also shown in underlined bold. Residues involved in the CsA binding site are in bold italics. This figure is done by Miss Jacqueline Dornan in the laboratory.

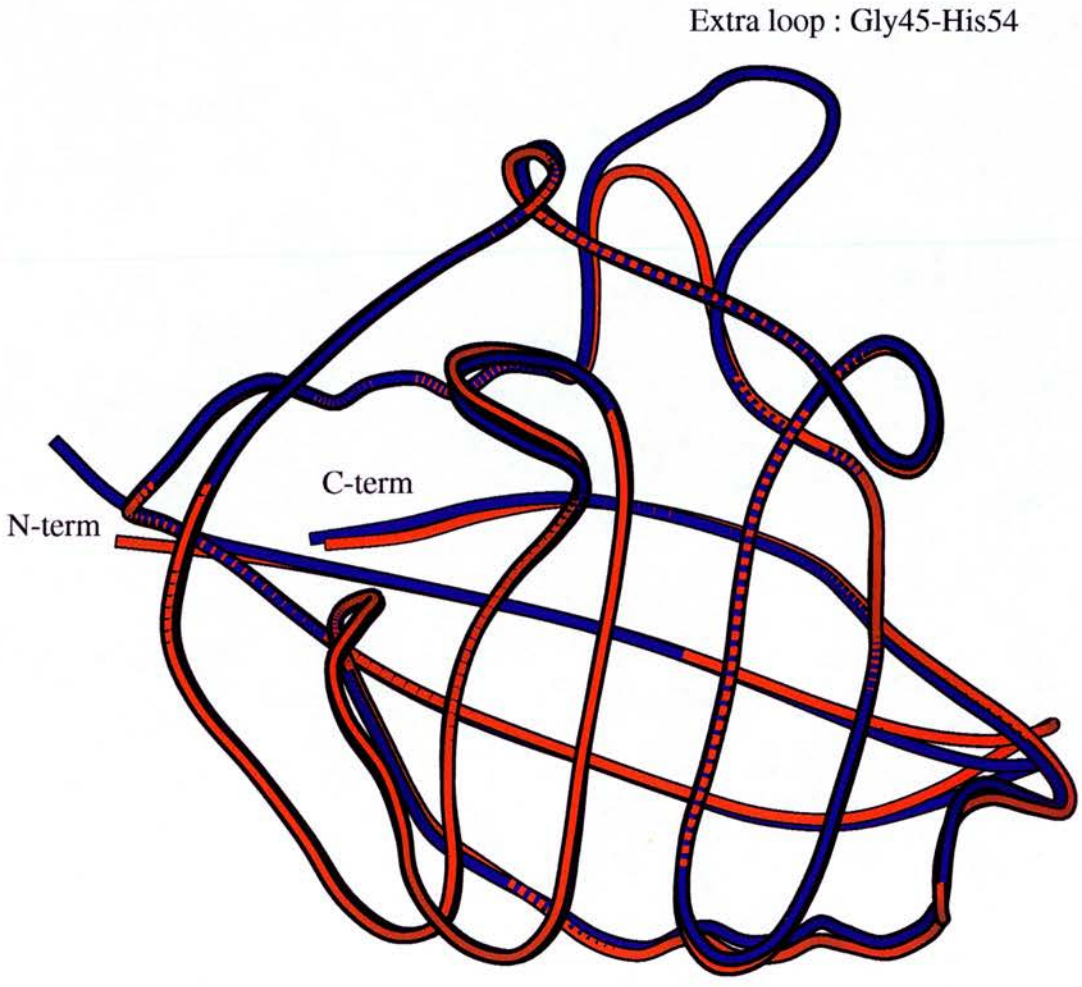


Figure 5-7.Diagram of the superimposition of the structures of hCYPA (red) and CYP3 (purple), produced by MOLSCRIPT (Kraulis, 1991).

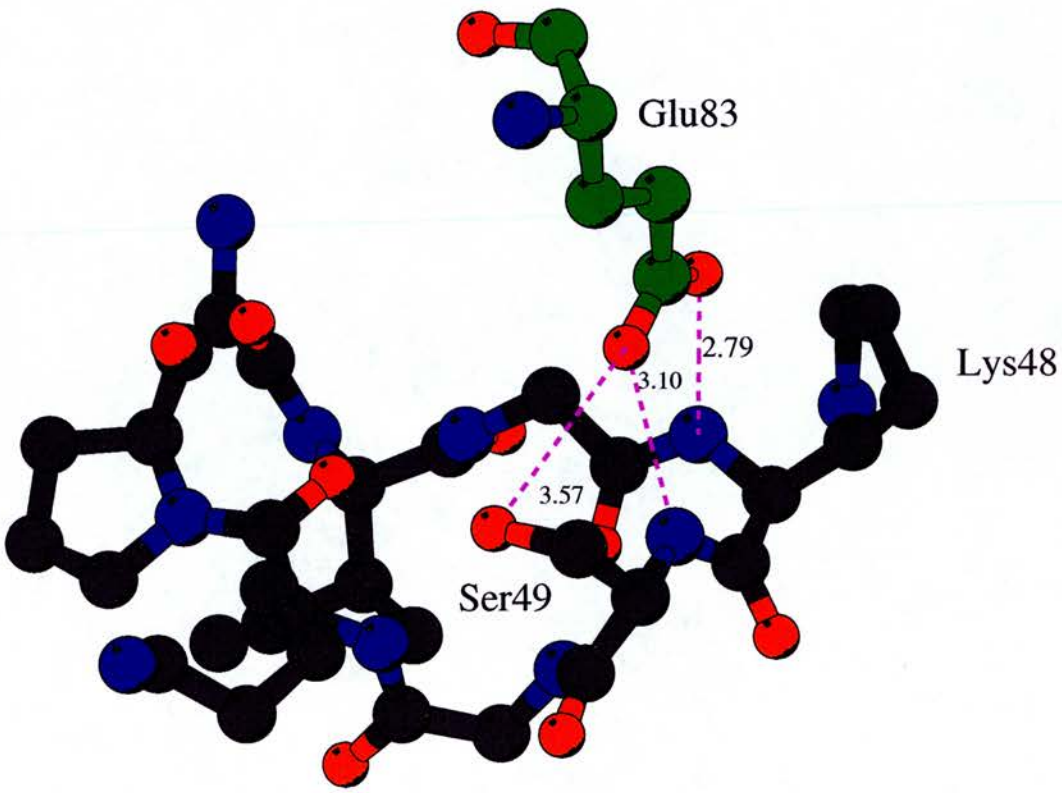


Figure 5-8.Diagram of the additional loop held by Glu83. Three hydrogen bonds formed by the side chain of Glu83 to Lys48 and Ser49 are shown as dashed lines. The numbers represent the hydrogen bond distances.

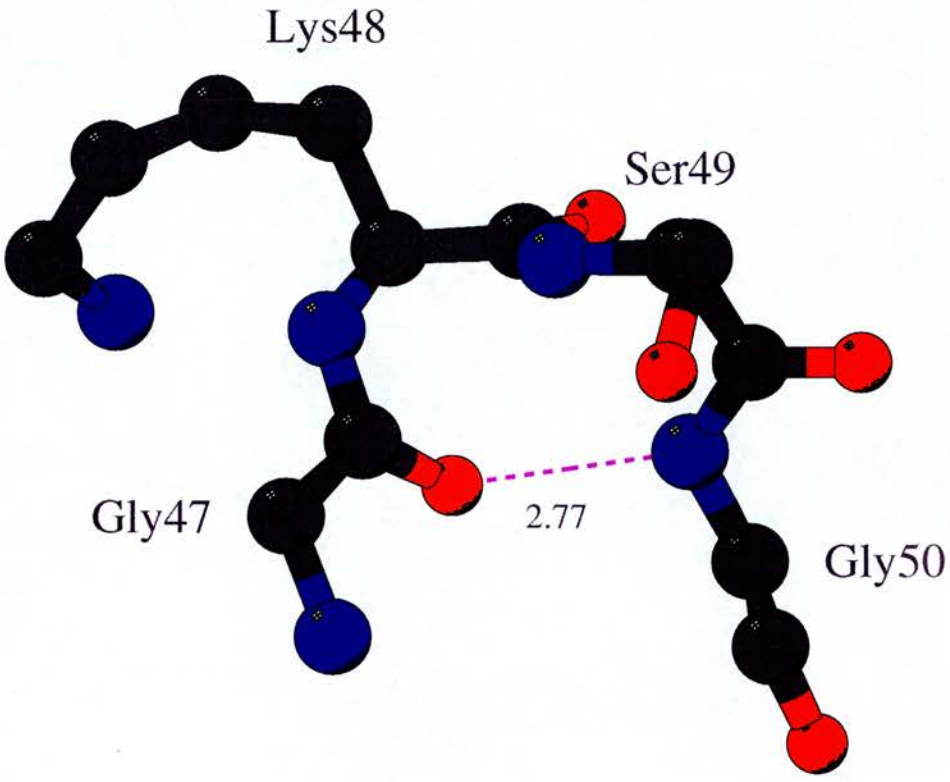


Figure 5-9.Diagram of the beta-turn found in the additional loop of CYP3. The hydrogen bond formed by the oxygen of Gly47 to the nitrogen of Gly50 with a bond distance 2.77A is shown as dashed lines.

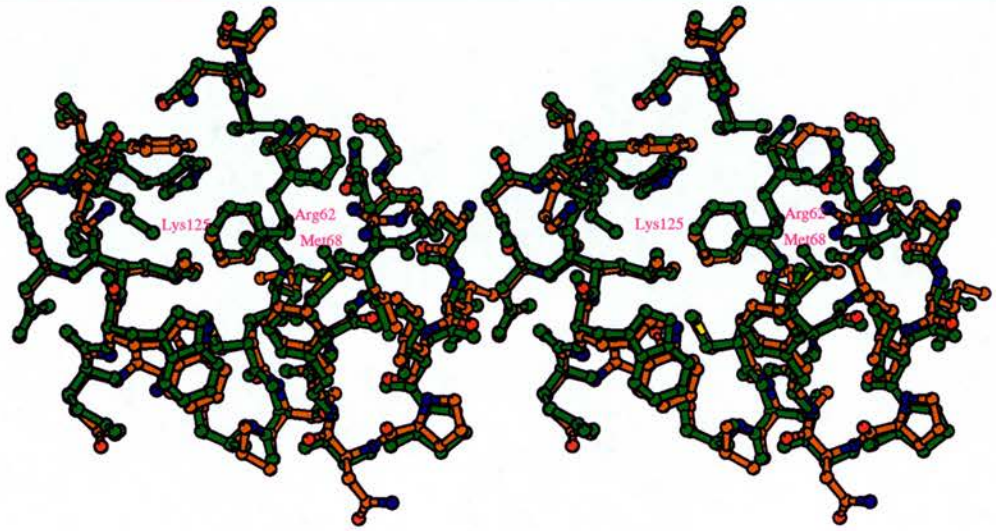


Figure 5-10.Stereodiagram of the superimposition of the binding sites of CYP3 and hCYPA (green). The labelled residues represent the residues that adopt the different conformations in these two structures.

5.2 The structure of Ala-Pro bound to CYP3

5.2.1 Refinement

The CYP3 native structure was used as a starting model to refine the structure of 120mM Ala-Pro bound to CYP3 (120mM Ala-Pro/CYP3). Positional and B-factor refinement were performed using SHELX97. After 10 cycles of positional and B-factor refinement, H_2O molecules were added using SHELXWAT. A (1Fo-1Fc) map was calculated using SHELXPRO and displayed with the program O. There is a clear density map within the expected binding site (See Figure 5-11).

The model of Ala-Pro was taken from the well-refined structure of Ala-Pro bound to hCYPA (PDB ID 1CYH) and built into the density with O. Also, from the density map, it is obvious Arg62 moves when Ala-Pro enters the binding site. Arg62 was repositioned into the density as well. Repeated cycles of positional and B-factor refinement were performed followed by the addition of H_2O molecules. The R-factor and R_{free} of each refinement stage are summarised in Table 5-C.

The final R-factor is 18.38% and R_{free} is 24.48% to 1.9Å resolution.

Table 5-C. The refinement process of the structure of Ala-Pro bound to CYP3

Refinement step	Resolution range(Å)	R (%)	R_{free} (%)
Positional and B-factor refinement	10-1.9	29.21	33.41
Adding 90 H_2O	10-1.9	23.43	29.00
Positional and B-factor refinement after building Ala-Pro and moving Arg62 into density	10-1.9	22.52	28.01
Adding another 160 H_2O	10-1.9	18.38	24.48

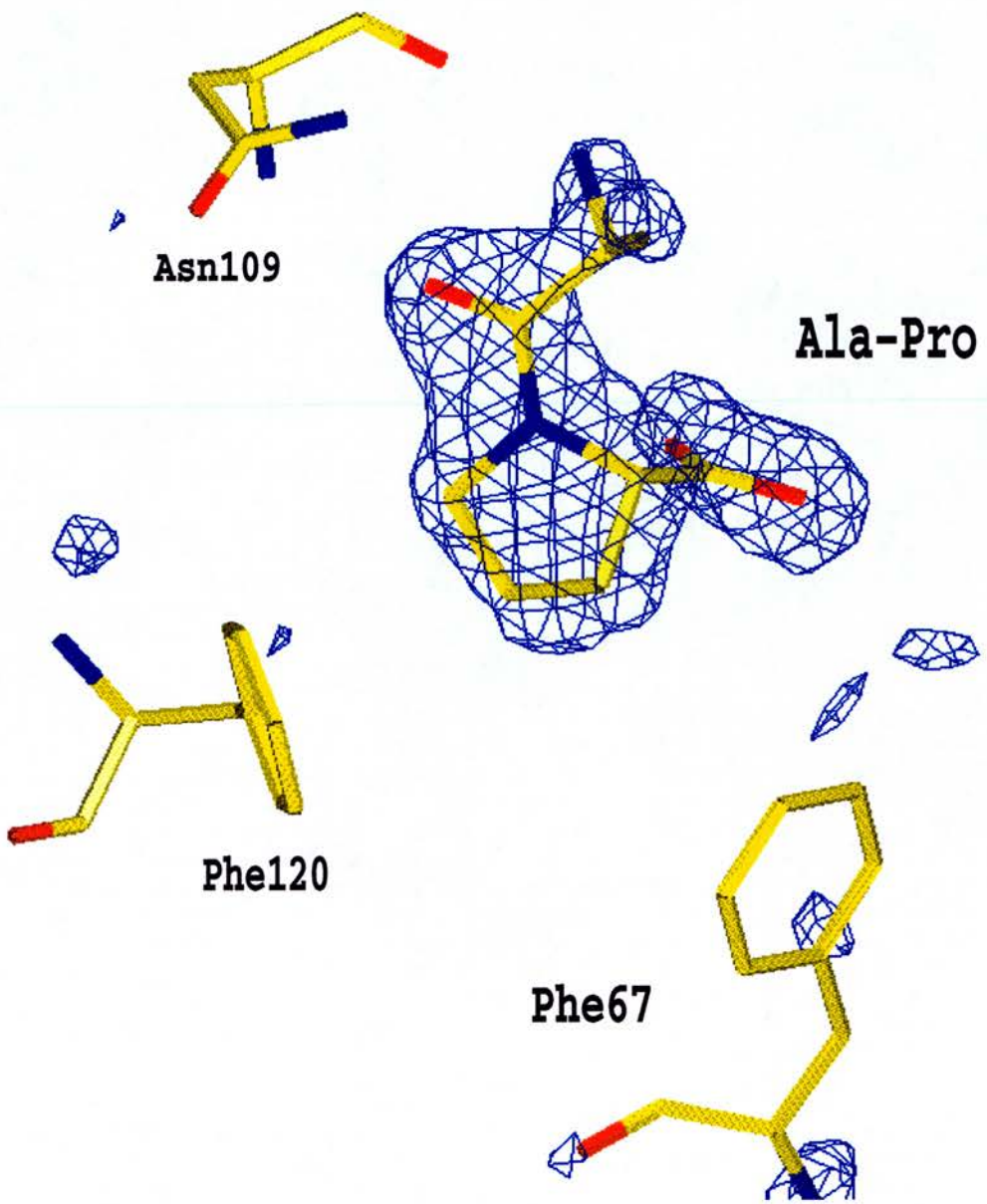


Figure 5–11.The (Fo–Fc) electron density map around Ala–Pro with the refined structure superimposed. The map, contoured at 1.5 sigma, was calculated by SHELXPRO (Sheldrick and Schneider, 1997) after the addition of water molecules in SHELX and displayed in O (Jones et al., 1991).

The final rms deviations of bond length and angle distance are 0.030Å and 0.028 Å, respectively. The average B-factor for all protein atoms is 32.13 $Å^2$. The stereochemical quality of the complexed structures assessed by PROCHECK is very similar to that of the native structure. No residues are in the disallowed regions. Two residues, Phe67 and Glu88, are in the generously allowed regions as they are in the native structure.

5.2.2 Binding site

The binding site is located outside the β-barrel and on the surface of the protein, which is similar to that of hCYPA as mentioned in section 1.2.4.

The side chain of proline makes contact with the hydrophobic pocket, which is composed of Phe67, Met68, Phe120, Leu129, His133. Specifically, Phe120 makes the most and strongest contact with the proline ring while the other residues only contribute one to two hydrophobic contacts to Ala-Pro. In addition, there are seven strong hydrogen bonds involved in the ligand-protein binding. The N-terminal nitrogen (N1) of Ala-Pro forms three hydrogen bonds with the carbonyl oxygen of Asn109 and two H_2O molecules while carbonyl oxygen (O1) of Ala-Pro forms one hydrogen bond with the NH nitrogen of Asn109. In addition, the two C-terminal oxygens form three hydrogen bonds with Arg62 and Gln70 (See Figure 5-12). The interatomic distances between Ala-Pro and CYP3 are summarised in Table 5-D.

The binding of Ala-Pro does not change the structure of the protein significantly. The rms deviation value of backbone atoms of 170 residues between the native and the complexed structure is 0.316Å. All residues around the binding site with the exception of Arg62 adopt an identical conformation to the native structure. In the native structure, three H_2O molecules occupy the binding position of Ala-Pro and the guanidino group of Arg62 adopts the conformation toward the binding site. In the structure of Ala-Pro bound to CYP3, when Ala-Pro approaches the binding site, the guanidino group of Arg62 has to move away to accommodate the ligand and consequently displaces the H_2O near Arg62 (see Figure 5-13)

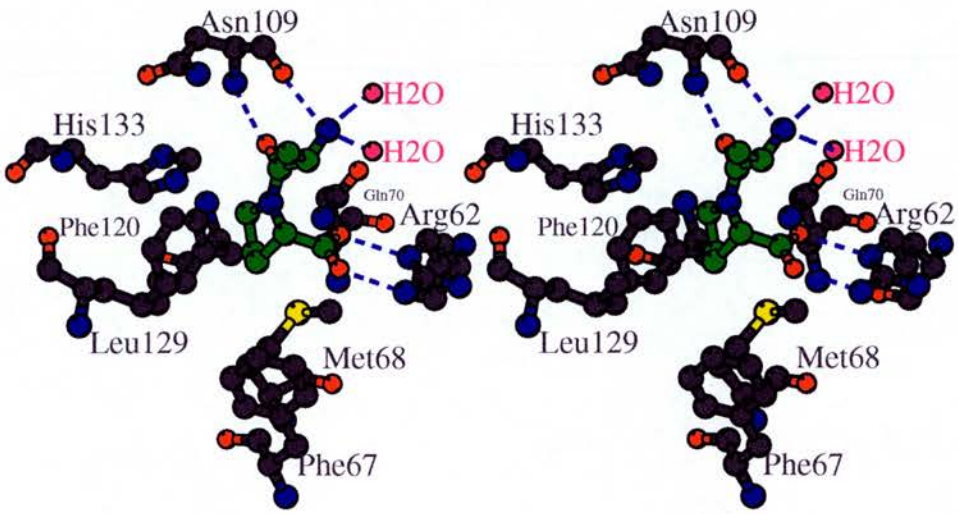
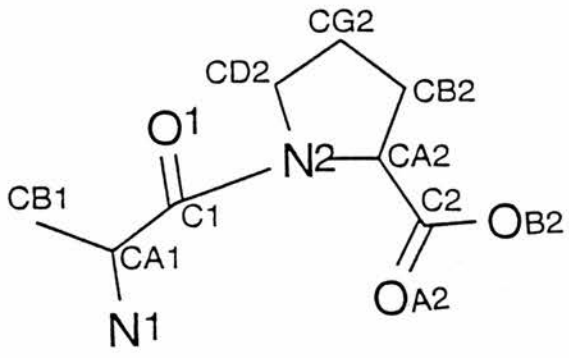


Figure 5-12.Stereodiagram of the binding site of CYP3 with bound Ala-Pro, produced by MOLSCRIPT (Kraulis, 1991). THe carbon atoms of Ala-Pro are shown in green. The residues formed the hydrophobic pocket and hydrogen bonded to Ala-Pro are labelled. Some of the hydrogen bonds are marked by broken lines.

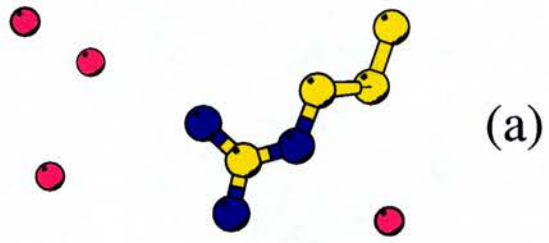
Table 5-D. (a) Chemical structure of Ala-Pro



ALA-PRO

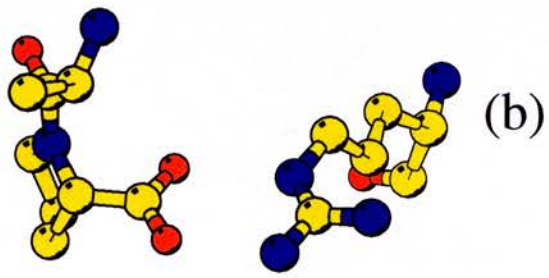
Table 5-D. (b) The interatomic distances between CYP3 and Ala-Pro

	Ala-Pro atom	CYP3 atom	Distance
Hydrophobic interactions	CD2	CE1/His133	3.68
	C1	CE1/His133	3.48
	CD2	CE1/Phe120	3.29
	CG2	CE1/Phe120	3.68
	CD2	CD1/Phe120	3.38
	CG2	CD1/Phe120	3.65
	CG2	CG/Phe120	3.88
	CB2	CZ/Phe67	3.66
	CB2	CD1/Leu129	4.00
	CG2	CE/Met68	3.91
Hydrogen bond interactions	N1	O/Asn109	2.73
	N1	HOH (no.363)	2.84
	N1	HOH (no.318)	2.69
	O1	N/Asn109	2.63
	OA2	NE2/Gln70	2.84
	OB2	NH2/Arg62	2.64
	OA2	NE/Arg62	2.77



(a)

H2O Arg62(N)



(b)

Ala-Pro Arg62(L)

Figure 5-13.(a) Diagram of the conformation of Arg62 (Arg(N)) when Ala-Pro is absent. The binding site is occupied by water molecules (magenta) when Ala-Pro is absent.(b) Diagram of the conformation of Arg62 (Arg(L)) when Ala-Pro binds.

5.2.3 Crystal dynamic study

Native crystals were soaked in solutions with different concentrations of Ala-Pro. The complexed structures with different ligand concentrations were refined using the same procedure as discussed above. The highest resolution of these structures varied from 1.9Å to 2.1Å and the final R-factor varied from 0.176 to 0.225. The quality of these structures is very similar while rms deviations of bond lengths are 0.031Å, the rms deviations of angle distance are around 0.026Å and the mean B-factor is around 32 $Å^2$. The stereochemical quality of all the complexed structures as assessed by PROCHECK is very similar to that of the native structure. No residues are in the disallowed regions. Two residues, Phe67 and Glu88, are in the generously allowed regions. The final refinement results for the complexed structures with different ligand concentrations are summarised in Table 5-E.

Table 5-E. Summary of the refinement statistics of the structures of CYP3 complexed with different concentrations of Ala-Pro

Summary of the refinement statistics	**1.2mM Ala-Pro**	**6mM Ala-Pro**	**18mM Ala-Pro**
Resolution range (Å)	10-2.1	10-2.0	10-1.9
R_{final} (F>4σ)	17.65% (16.52%)	18.09% (17.10%)	20.93% (19.45%)
R_{free}[1] (F>4σ)	25.93% (24.96%)	26.54% (24.43%)	29.61% (27.21%)
Rms deviations from ideal geometry			
Bond lengths (Å)	0.030	0.030	0.030
Angle distance (Å)	0.027	0.027	0.027
mean B-factor ($Å^2$)	33.57	31.82	32.67

Summary of the refinement statistics	**36mM Ala-Pro**	**60mM Ala-Pro**	**120mM Ala-Pro**
Resolution range (Å)	10-1.9	10-2.0	10-1.9
R_{final} (F>4σ)	19.50% (18.33%)	22.58% (19.96%)	18.38% (17.3%)
R_{free}(F>4σ)	26.85% (24.63%)	29.21% (26.41%)	24.48% (23.77%)
Rms deviations from ideal geometry			
Bond lengths (Å)	0.031	0.030	0.030
Angle distance (Å)	0.027	0.025	0.028
mean B-factor ($Å^2$)	31.14	35.76	32.13

1 R_{free} is calculated with 5% of the data omitted from refinement for cross-validation

After positional and B-factor refinement, fitting Ala-Pro and Arg62 into the density and adding H_2O, the occupancy of Ala-Pro was refined using SHELX97. The coordinates of $Arg62_{(N)}$, $Arg62_{(L)}$, three H_2O molecules in the binding site, one H_2O molecule near $Arg62_{(N)}$ and Ala-Pro were fixed during occupancy refinement. $Arg62_{(N)}$ represents the Arg62 conformation without Ala-Pro binding to CYP3 while $Arg62_{(L)}$ represents the Arg62 conformation with Ala-Pro binding to CYP3. The B-factor of $Arg62_{(N)}$, 3 H_2O molecules in the binding site and one H_2O near $Arg62_{(N)}$ were fixed to the same B-factor as that of the native structure. The B-factor of $Arg62_{(L)}$ and Ala-Pro were fixed to the value as the same as the B-factor of those obtained from the structure of 120mM Ala-Pro bound to CYP3. Since the binding site is occupied by either ligand or H_2O, the sum of the occupancies of Ala-Pro and H_2O is 1. In addition, because Arg62 adopts either the native or binding conformation, the sum of these two conformations is 1 as well. Further, the occupancy of Ala-Pro is the same as that of $Arg62_{(L)}$ since the Arg62 adopts the binding conformation only when the binding site is occupied by Ala-Pro. Moreover, the occupancy of H_2O is the same as that of $Arg62_{(N)}$ since the Arg62 adopts the native conformation only when the binding site is occupied by H_2O. The restraints applied in the occupancy refinement are summarised below:

$$
\begin{array}{lllll}
Q_{H2O\,(a)} & = & Q_{H2O\,(b)} & & \\
Q_{H2O\,(a)} & + & Q_{Ala\text{-}Pro} & = & 1 \\
Q_{Arg\,62(N)} & + & Q_{Arg62(L)} & = & 1 \\
Q_{Arg62\,(L)} & = & Q_{Ala\text{-}Pro} & & \\
Q_{Arg62\,(N)} & = & Q_{H2O\,(a)} & = & Q_{H2O\,(b)}
\end{array}
$$

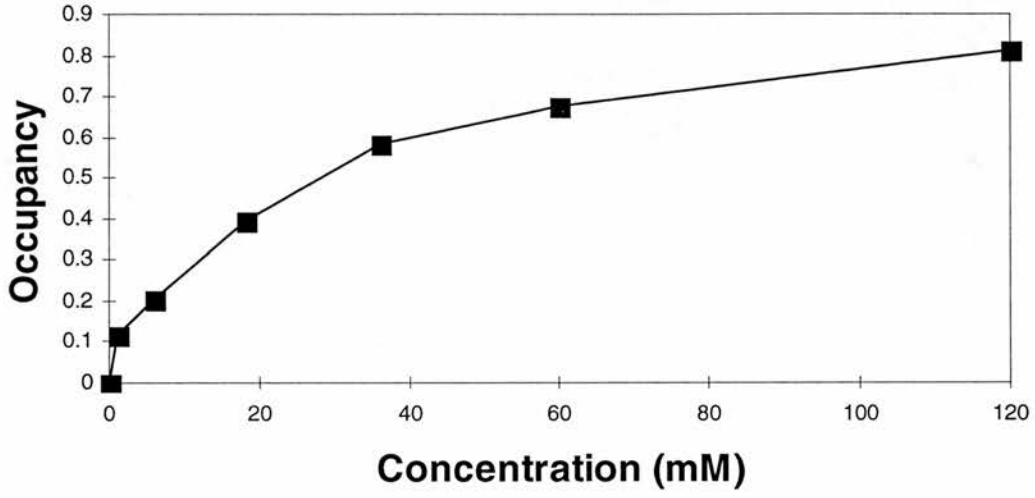
where

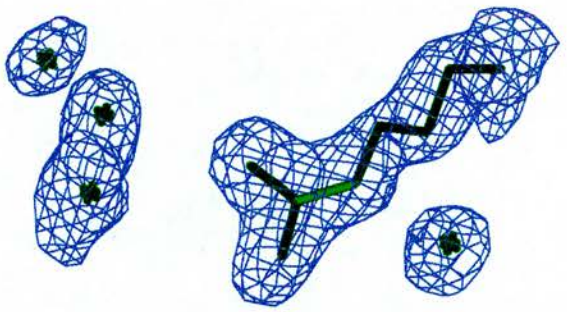
$Q_{H2O\,(a)}$: occupancy of H_2O in the binding site while no Ala-Pro binding
$Q_{H2O\,(b)}$: occupancy of H_2O near NE/Arg62 while no Ala-Pro binding
$Q_{Arg62\,(N)}$: occupancy of Arg62 conformation while no Ala-Pro binding
$Q_{Arg62\,(L)}$: occupancy of Arg62 conformation while Ala-Pro binding
$Q_{Ala\text{-}Pro}$: occupancy of Ala-Pro

The occupancy of Ala-Pro at different soaking concentrations is shown in Table 5-F. The (2Fo-1Fc) electron density maps around the binding site of CYP3 complexed with different concentrations of Ala-Pro are shown in Figure 5-14.

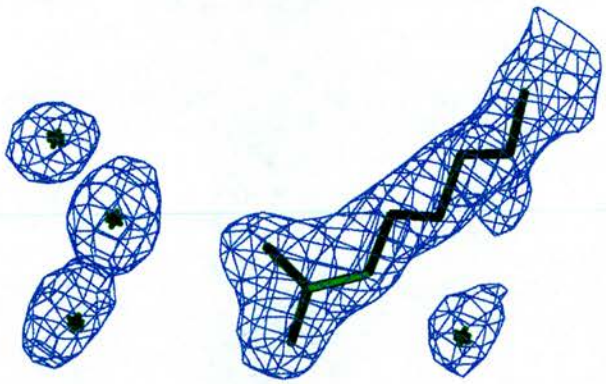
Table 5-F. The occupancy of Ala-Pro at different soaking concentrations

Soaking concentration of Ala-Pro (mM)	Occupancy of Ala-Pro
0	0
1.2	0.119
6	0.206
18	0.396
36	0.586
60	0.678
120	0.812

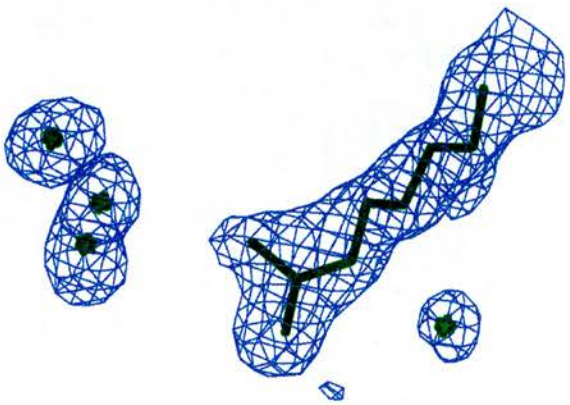




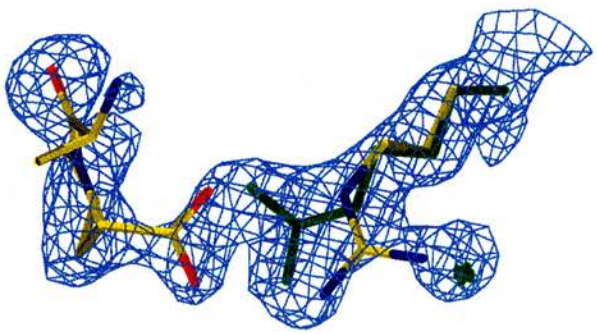
(a)



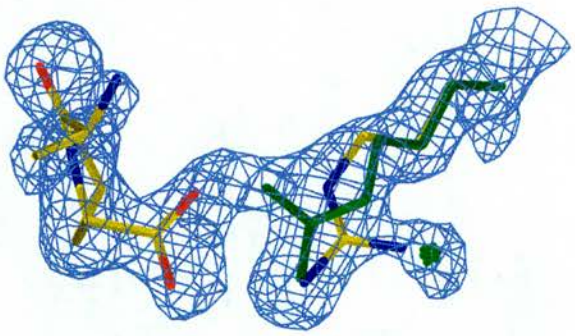
(b)



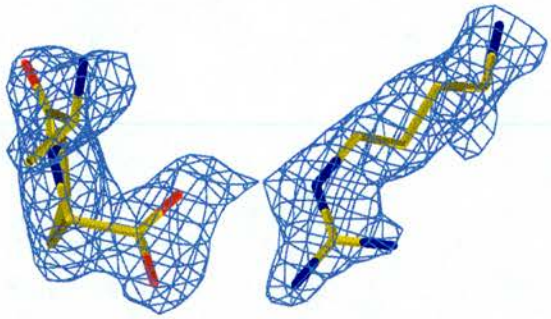
(c)



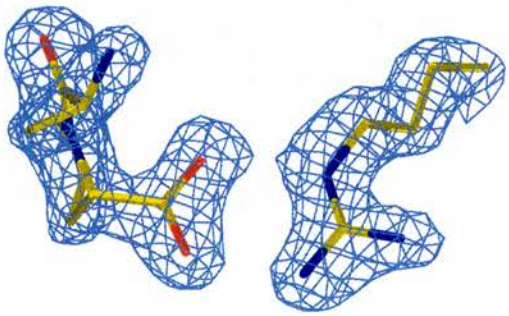
(d)



(e)



(f)



(g)

Figure 5–14.The (2Fo–Fc) electron density map around the binding site of CYP3 complexed with different concentrations of Ala–Pro. All maps were calculated by SHELXPRO (Sheldrick and Schneider, 1997) and displayed in O (Jones et al., 1991). (a) The density map around the binding site of native CYP3 structure, contoured at 1.5 sigma. Water molecules are seen in the binding site when Ala–Pro is absent and Arg62 adopts the conformation Arg62(N). (b) The density map, contoured at 1.5 sigma, around the binding site of the structure of CYP3 complexed with 1.2mM Ala–Pro. The map is very similar to (a) and the electron density of Ala–Pro is too weak to be seen in the map.(c) The density map, contoured at 1.5 sigma, around the binding site of the structure of CYP3 complexed with 6mM Ala–Pro. Similar to (b),the electron density of Ala–Pro is too weak to be seen in the map. (d) The density map, contoured at 1.5 sigma, around the binding site of the structure of CYP3 complexed with 18mM Ala–Pro. The density of Ala–Pro starts to show up although the shape is not well–defined. It also could be seen from the density map that Arg62 adopts two conformations (Arg62(N) and Arg62(L))(e) The density map, contoured at 1.5 sigma, around the binding site of the structure of CYP3 complexed with 36mM Ala–Pro. The density of Ala–Pro is very clear. Similar to (d),Arg62 adopts two conformations (Arg62(N) and Arg62(L))(f) The density map, contoured at 1.7 sigma, around the binding site of the structure of CYP3 complexed with 60mM Ala–Pro. The density of Ala–Pro is very clear. From the density map, it shows that Arg62 adopts the conformation (Arg62(L)). (g) The density map, contoured at 1.7 sigma, around the binding site of the structure of CYP3 complexed with 120mM Ala–Pro. Similar to (f), the density of Ala–Pro is very clear and Arg62 adopts the conformation (Arg62(L)).

The (2Fo-1Fc) maps around the binding site of 1.2mM and 6mM Ala-Pro/CYP3 complexed structure are very similar to that of the native. This indicates the electron density of Ala-Pro is too weak to be seen in the map and only H_2O molecules are seen in the binding site, as the occupancy of Ala-Pro is very low in both structures. With 18mM Ala-Pro, the density of Ala-Pro starts to show up although the shape is not well-defined, especially the CA1, CB1 and N1 atoms of Ala. Also, the density map shows the mixed-conformations adopted by Arg62 ($Arg62_{(N)}$ and $Arg62_{(L)}$) and indicates the occurrence of the $Arg62_{(N)}$ conformation is close to that of the $Arg62_{(L)}$ conformation. As mentioned before, the sum of the occupancy of $Arg62_{(N)}$ and that of $Arg62_{(L)}$ is 1. If the occurrence of these two conformations is equal, the theoretical occupancy would be 0.5 for each. As might be expected, if the density map shows the mixed-conformations adopted by Arg62, the occupancy of $Arg62_{(L)}$ and $Arg62_{(N)}$ must be near 0.5. The density map is consistent with the occupancy calculated from SHELX97 while the occupancy of Ala-Pro, equal to $Arg62_{(L)}$, is 0.396. in this structure. When the concentration of Ala-Pro increases, the density map becomes more interpretable. As shown in the density map of 36mM Ala-Pro/CYP3 complexed structure, the map of Ala-Pro is very clear and the model fits perfectly into the density. As for Arg62, the density map still shows the mixed conformations of Arg62. The occupancy of Ala-Pro is greater than 0.5. While in the 60mM and 120mM Ala-Pro/CYP3 complexed structures, Ala-Pro fully occupies the binding site and the liganded conformation of Arg62 dominates in the density map as shown in Figure 5-14 (f) and (g). The occupancies of Ala-Pro in these two complexed structures are 0.678 and 0.812, respectfully.

5.2.4 Calculation of the dissociation constant *"in crystallo"*

$$[\ PL\ ] \rightleftharpoons [\ P\ ] + [\ L\ ]$$

$$K_{dc} = [\ P\ ][\ L\ ] / [\ PL\ ]$$

Because

$$[\ P_{total}\ ] = [\ PL\ ] + [\ P\ ]$$
$$[\ PL\ ] = [\ P_{total}\ ]\ Q$$
$$[\ P\ ] = [\ P_{total}\ ]\ (1 - Q)$$

Thus,

$$K_{dc} = \{\ [\ P_{total}\ ]\ (1 - Q)\ \}\ [\ L\ ] /\ [\ P_{total}\ ]\ Q$$
$$= (1-Q) * [\ L\ ] / Q$$

where

K_{dc} : dissociation constant *"in crystallo"*
$[\ P\ ]$: concentration of protein without ligand
$[\ L\ ]$: concentration of ligand
$[\ PL\ ]$: concentration of protein with ligand
$[\ P_{total}\ ]$: total concentration of protein
Q : occupancy of ligand

It was assumed that the concentration of ligand is constant throughout the soaking experiment. This is a reasonable assumption as the concentration of ligand is much greater than that of protein, around 1μM.

Re-arranging the equation as follows:

$$K_{dc} = (1-Q) * [\ L\ ]/ Q$$
$$\Longrightarrow 1/\{(1/Q)-1\} = 1/ K_{dc}\ [L]$$

$1/ K_{dc}$ is the proportionality constant for the plot of $1/\{(1/Q)-1\}$ versus [L].

This result is presented in Figure 5-15.

The experimental data fit the theoretical calibration curve with a high R^2, 0.9981.

$1/k_{dc}$ is 0.036 from the calibration calculation. Therefore, the dissociation constant *"in crystallo"*, K_{dc}, is 27.7mM.

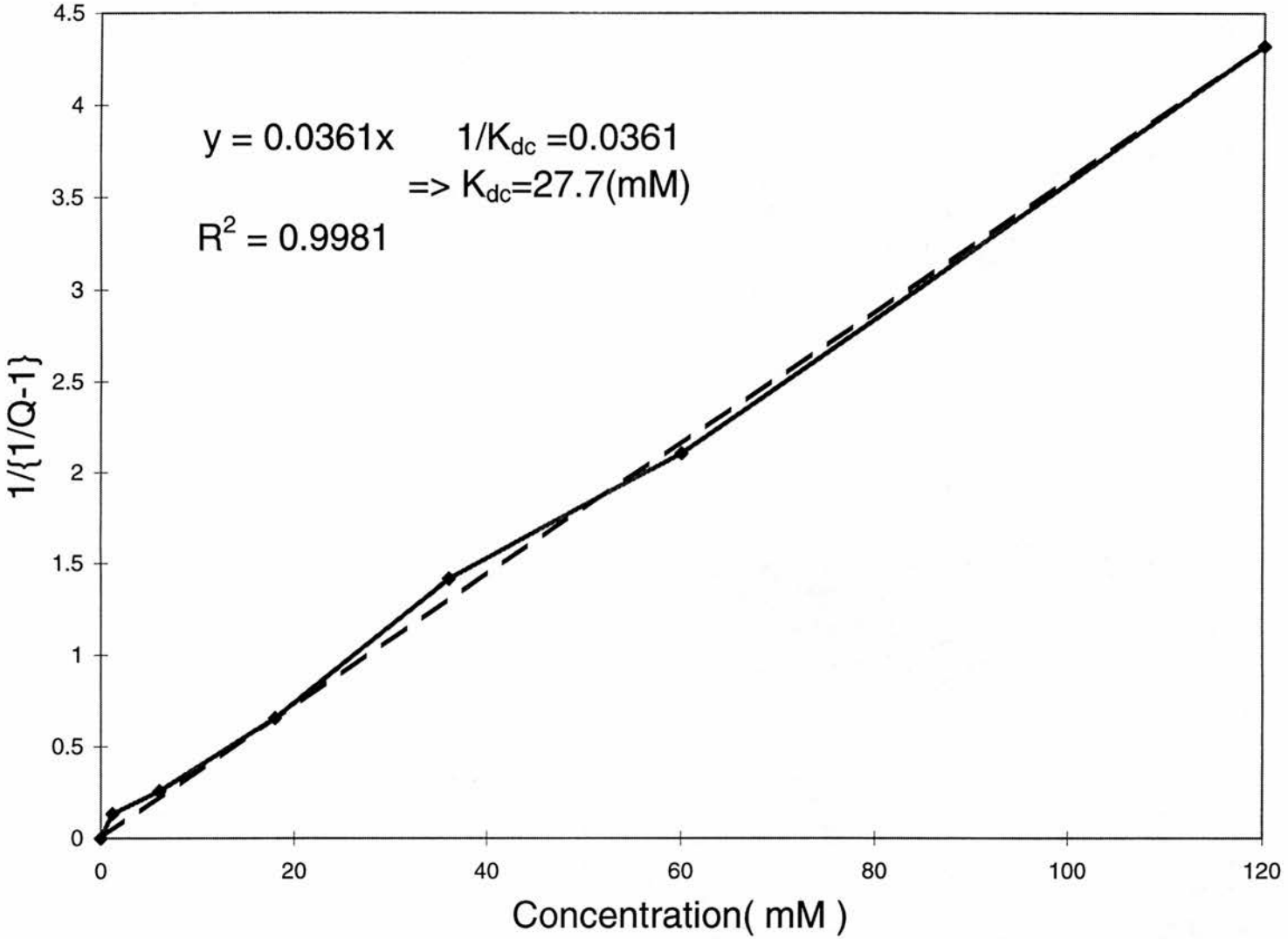


Figure 5-15. The plot of 1/{(1/Q)-1} versus [L]. The solid line represents the experiment data while the dash line represents the calibration curve. K_{dc}, obtained from the reverse of the proportionality constant of the plot, is 27.7mM.

5.2.5 Discussion

According to the preliminary solution study, the dissociation constant of Ala-Pro binding to CYP3 determined by PPIase assay is 31.23 mM (Dornan, 1999a). As can be seen, the dissociation constant obtained by x-ray crystallography is comparable with one obtained by solution studies.

The different values obtained from these two methods probably result from the different experimental systems as one is in solution while the other is in crystal. Moreover, the imprecise nature of the occupancy determination by the refinement program, partially due to the high correlation that exist between occupancy and B-factor may lead to the inaccuracy of the dissociation constant. In order to overcome this problem, the B-factors are fixed during occupancy refinement as mentioned in section 5.2.3. However, this fixation is valid only when the quality of each data is similar as the B-factors decrease if the quality of data is better while the B-factors increase if the quality of data is worse. In the model of Ala-Pro bound to CYP3, the quality of each data is similar as the resolution and R-factor vary in a small range between these data. Recently, the "frozen and unfrozen" method (Yeh & Hol, 1998) as mentioned in section 2.4.1 enable to collect more than one data set from the same crystal. If this method can be applied successfully, it will improve the determination of occupancy as the quality of each data will be expected to be very similar because of the same crystal. In the current experiment, the CYP3 crystals cracked when unfrozen. Therefore, different cryoprotectant solution can be tried in order to achieve the goal.

Nevertheless, the structure of Ala-Pro bound to CYP3 provides a good model to study the dynamic behaviour of ligand binding *"in crystallo"*. Firstly, the data and the protein structures are good enough to present the successive change of the density map as the result of the increase of the ligand concentration, for example, the gradual appearance of Ala-Pro, disappearance of the water molecules and the movement of Arg62. Secondly, the occupancy of Ala-Pro can be determined more precisely as more than one restraints are applied in the occupancy refinement (see section 5.2.3). More restraints can provide more information to calculate the occupancy of ligand

and consequently give rise to the more accurate determination of occupancy. As can be seen, the occupancy of Ala-Pro calculated by SHELX are reasonable as the occupancy increases when the concentration of Ala-Pro increases and the R^2 of the calibration curve obtained from the plot of 1/ { (1/Q)-1) versus [L] is very high (R^2 = 0.9981), which shows the experimental data fit the theoretical value very well.

6. Conclusion

This chapter will summarise the results of the work that came out of this thesis and will bring together the various components previously discussed in earlier sections, providing an overview of the work undertaken. Finally, some suggestions for future studies will be given at the end of the chapter.

6.1 BLG

The structure of BLG co-crystallised with palmitate at pH 7.5 has been solved by molecular replacement using a monomer of the dimeric BLG in lattice X form as the search model. The complexed structure has been refined to 2.5Å with R=0.204 (R_{free}=0.240) for 6888 reflections. It has been shown that the palmitate binds in the central cavity of the molecule, formed by the eight antiparallel β-strands. The hydrophobic tail of palmitate stretches into the interior of the calyx while the carboxyl group of palmitate make hydrogen bond contacts with both Lys60 and Lys69 at the open end of the barrel. In addition to the hydrogen bonds, hydrophobic interactions play an important role in the binding of palmitate to BLG. There are a total of 29 hydrophobic contacts with the 13 residues in the hydrophobic cavity over a distance of less than 4Å. In particular, the side chain of Phe105 faces the tail of the palmitate and contributes most to the hydrophobic interaction.

BLG does not undergo a dramatic change in conformation upon palmitate binding. All residues in the hydrophobic cavity with the exception of Met107 remain in the native conformation. The side chain of Met107 adopts an alternative rotamer to accommodate the ligand. In order to fit inside the calyx, the conformation of palmitate bends slightly from a fully extended all-trans conformation. There is a kink at C6 of palmitate associated with the movement of the side chain of Met107. Careful examination of the B-factors of the palmitate shows the carbon atoms of the tail (C_8-C_{16}) have lower B-factors whereas those near the opening of the calyx have much higher B-factors. The difference in B-factors can be explained by the distinct interactions between ligand and protein. Firstly, the tail of palmitate makes strong hydrophobic contacts with the surrounding residues and is thus stabilised by

hydrophobic interactions, resulting in the lower B-factors seen. Secondly, the significant increase of B-factors of the atoms near the entrance of the calyx is presumably reflecting the larger volume of the cavity and hence the less specific nature of the binding in the region of the highly flexible EF and GH loops.

By comparing the structure of palmitate-bound BLG at pH 7.5 with the structure of the native lattice X at pH 6.5, it has been shown the EF loop moves when the pH is increased from pH 6.5 to pH 7.5. In the lattice X structure, the EF loop is folded over the entrance thus hindering the access to the calyx. However, in the structure of palmitate binding to BLG, the EF loop is folded back to allow palmitate access to the binding site. Therefore, it can be concluded that the movement of EF loop is essential to open up the entrance to the calyx. The pH-dependant movement of the EF loop is also responsible for the well-characterised Tanford transition (Tanford *et al.*, 1959). Qin and co-workers have refined the structures of the lattice Z form at 3 distinct pH values (6.2, 7.1, 8.2) showing that there is a distinct movement of a loop as the pH is raised (Qin *et al.*, 1998a). This movement uncovers a buried carboxyl group, observed during titration by Tanford *et al* (1959). This anomalous carboxyl group has been confirmed as the side chain of Glu89 by a recent structural study (Qin *et al.*, 1998). The feature of the EF loop serving as a portal loop is somehow similar to that of the dynamic portal hypothesis (Hodsdon and Cistola, 1997b) in fatty acid binding protein (FABP). In FABP, the access and binding of fatty acid is controlled by a flexible region that consists of several disordered loops. In BLG, discrete arrangements of the 'portal' EF loop exist and some nearby loops (AB, GH and CD loops) at the open end of the calyx are very flexible to allow the entry of the ligand without obstruction.

The structure of BLG complexed with a fatty acid analogue, 12-bromododecanoic acid, has been solved independently by Qin *et al* (1998b). The complexed structure shows 12-bromododecanoic acid binds to BLG in a manner similar to palmitate with the tail extended to the bottom of the calyx and the head located at the entrance of the calyx. Comparison of the structure of 12-bromododecanoic acid-bound BLG and that of palmitate-bound BLG reveals the major difference is that palmitate makes more hydrophobic contacts with BLG. This observation of the greater degree of

hydrophobic contact of palmitate with BLG is consistent with the result seen by solution studies using fluorescence which shows that the binding strength of palmitic acid is stronger than that of dodecanoic acid (lauric acid).

The binding site in BLG is fully extended, but there is space for longer fatty acid molecules such as stearate and oleate to be accommodated within the calyx, with the carboxyl group making the same interactions with Lys 60 and Lys 69. The association constants for both acids are similar.

At this stage, little can be said about the existence of a second binding site. It is surprising to find palmitate in the calyx, since the report of Narayan & Berliner (1997) establishes fatty acid binding at a site which is not perturbed by retinol and the report of Cho *et al.* (1994) implicates Lys 69 with retinol binding. Furthermore, contrary to the findings of Narayan and Berliner, Puyol *et al* (1991) find that palmitate and retinol bind competitively to BLG, the former displacing the latter. Thus, the groove identified by Monaco *et al.* (Monaco *et al.*, 1987) on the outer surface of the protein has yet to be confirmed as a binding site for any ligand, despite several strands of circumstantial evidence that point to its existence (Sawyer *et al.*, 1998). Thus, to collect a suitable data from the retinol/BLG complex is necessary to reveal the retinol binding site.

6.2 CYP3

6.2.1 Native structure

The structure of cyclophilin 3 (CYP3) has been solved and refined to 1.8Å with R=0.208 (R_{free}=0.284). The overall structure of CYP3 contains a β-barrel formed by 8 antiparallel β-strands with 2 α-helices sitting on the top and bottom of the barrel. The inside of the barrel is a hydrophobic core formed by the side chains of Val6, Phe8, Ile20, Met22, Ile63, Ile69, Leu105, Phe119, Leu121, Ile165. These two α-helices block the entrance to the core and consequently hinder the entrance of ligands to the core. Moreover, inside the core, the hydrophobic residues pack tightly excluding the possibility of ligand binding. Therefore, no ligands or substrates are able to bind inside the β-barrel. Instead, ligands bind in a long and deep groove

located on the surface of the protein between one face of the β-barrel and the FG loop.

The structure of CYP3 is very similar to that of human cyclophilin A, the best-characterised member of cyclophilin family. The binding site, which is also the active site, of CYP-3 is almost identical to that of CYPA. All residues important in ligand binding are conserved, the only differences being that the side chains of Arg62, Met68 and Lys125 are slightly shifted. Therefore, it is not surprising to find that the catalytic efficiency of CYP3 is comparable with that of CYPA as the IC_{50} value of CYP3 determined by the inhibition of CsA is 16nM, which is in agreement with the value of hCYPA, 19nM. It is therefore distinctly possible that ligands would bind to CYP3 in a similar way to that seen with CYPA, and that the isomerase activity of CYP3 should resemble that of CYPA although this conclusion can not be confirmed until more experiments have been done.

Comparison of these two structures reveals the major difference is the additional loop found in CYP3. This additional loop, located above the active site consists of 7 inserted residues (48KSGKPLH54). Interestingly, this loop is held in place by a tight and specific network of hydrogen bonds from the side chain of Glu83, the carboxylate group of Glu83 making three strong hydrogen bonds with Lys48 and Ser49. The role of Glu83 is probably to lock the loop into a particular β-turn conformation, which in other structures is usually involved in protein-protein interactions or receptor binding. Several cyclophilin homologues, for example, *Brugia malayi* cyclophilin, human/bovine cyclophilin 40 and some plant cyclophilins, have also this additional loop. It is particularly common in the cyclophilin isoforms from the nematode *C.elegans*. Therefore, this inserted loop may be used as an important recognition feature of this subclass of cyclophilins.

All proteins with the additional loop have a conserved glutamate while a lysine in the homologous position is found in the cyclophilins without the additional loop. The high degree of homology of this additional loop and the conserved glutamate residue indicate both the loop structure and locking mechanism are likely to play an important role although the actual role of this specific loop is still unclear.

Another unusual feature of CYP3 is the unexpected presence of the reduced forms of Cys40 and Cys168. The two sulphur atoms of Cys40 and Cys168 are very close, separated by 5.38 Å, but fail to form the expected disulphide bridge. Modelling studies have suggested Cys40 and Cys168 would be able to form a disulphide bridge if the chi1 torsion angle of both residues is rotated to the appropriate angle. The biological roles these unusual cysteines play require further investigation although some possible functions have been suggested in this thesis.

6.2.2 The structure of Ala-Pro bound to CYP3

The structure of CYP3 complexed with 120mM Ala-Pro has been refined to 1.9Å with an R-factor of 18.38% (R_{free} = 24.48%). The density map around Ala-Pro has clearly shown that the ligand binds in a deep groove on the surface of the protein. Both hydrophobic and hydrogen bond interactions contribute to the ligand-protein binding. The side chain of proline occupies the hydrophobic pocket of the protein which consists of the following residues, Phe67, Met68, Phe120, Leu129 and His133, and makes strong contacts with the surrounding residues, especially Phe120. Additionally, seven strong hydrogen bonds are formed by Ala-Pro with residues Asn109, Arg62, Glu70 and also two water molecules in the binding site. CYP3 does not undergo a dramatic structural change upon Ala-Pro binding. All residues around the binding site, with the exception of Arg62, adopt the conserved conformations to the native structure. In the structure of Ala-Pro bound to CYP3, when Ala-Pro approaches the binding site, the guanidino group of Arg62 has to move away to accommodate the ligand and consequently displaces the H_2O near Arg62. Moreover, the structure of Ala-Pro bound to CYP3 was used as a model to study the dynamic feature *"in crystallo"*. A series of data from native crystals soaked in different concentration of Ala-Pro have been collected. The occupancy of Ala-Pro in each data has been determined using SHELX97. Once the occupancy of Ala-Pro has been determined, the dissociation constant *"in crystallo"* can be calculated with the application of the appropriate mathematic equations. The dissociation constant obtained by x-ray crystallography is 27.7mM. This value is comparable with the value obtained using the PPIase assay in solution.

6.3 Future work

(a)BLG:

The most difficult problem in the study of BLG with hydrophobic ligands by x-ray crystallography is how to introduce sufficient ligand into the protein without damaging the crystal as the ligands are usually insoluble in water. The co-crystallisation of BLG with palmitate has been successful in this thesis. Therefore, the protocol of co-crystallisation developed here could be applied to introduce other hydrophobic ligands into BLG.

Retinol will be the most interesting ligand in the future studies. Actually, a crystallographic study of the binding of retinol to BLG has been carried out by Mr. George Kontopidis in the laboratory. Crystals of retinol-bound BLG have been obtained by modifying the protocols mentioned in section 2.3.2. The structure of retinol bound to BLG has been refined using the protein structure of palmitate bound to BLG as a starting model with the model of palmitate omitted. From the difference map, it is clear that retinol binds inside the calyx. Once the refinement of the structure of BLG complexed with retinol is complete, this complexed structure will be compared with that of BLG complexed with palmitate. The similarity and difference of these two complexed structures may provide more information about the biological function of BLG.

In addition to retinol, other hydrophobic ligands as mentioned in section 1.1.6.3, phosphate lipids and some vitamin analogues, for examples, vitamin D and D2 could be examined by performing similar experiments to those described in this thesis in order to investigate the existence of the other binding site and most important, to understand the nature and the biological roles of BLG.

As for the future work of BLGY structure, it is clear that the type of buffer and precipitant affect the crystallisation in a subtle way. Full refinement, probably of data from crystals from several different batches, would be one way of trying to understand the subtle variations of structure that lead to poor refinement and variations in the achievable resolution in any given case.

(b)CYP3:

The structure of Ala-Pro bound to CYP3 is the first model system that has allowed determination of the dissociation constant *"in crystallo"* by x-ray crystallography. In future studies, more ligands, for example, CsA, tetrapeptides and other dipeptides, could be tested to investigate whether the structure or the size of the ligand affects the accuracy of the determination of the dissociation constant *"in crystallo"*. Moreover, an alternative crystal form of CYP3 could be crystallised. Once the crystals in a different crystal form are obtained, the same experiments on CYP3 as performed in this thesis could be repeated to obtain the dissociation constant of Ala-Pro bound to CYP3 in the alternative crystal form. The comparison of these two dissociation constants determined from the different crystal forms will reveal whether the crystal packing affects the ligand binding behaviour *"in crystallo"*. Actually, the other crystal form of CYP3 has been obtained at 4°C by the hanging drop vapour diffusion method. The well solution contains 26% MPEG5000 as precipitant and 0.1M sodium citrate adjusted with citrate acid to pH 6.0. For a hanging drop, 2µl 21mg/ml CYP3 in 20mMHEPES/0.5MNaCl buffer was added to 2µl well solution. The data have been collected and processed by Dr. Paul Taylor in the laboratory. These crystals are orthorhombic with unit cell dimensions of a=58.9Å, b=67.5Å, c=76.4Å. However, no further studies of this crystal form were carried out as they are very anisotropic. Therefore, the future work will be involved in modifying the crystallisation conditions in order to improve the quality of the crystal in the orthorhombic form.

Finally, other model systems of different proteins complexed with different ligands could be examined to provide more information to insight into the dynamic interactions between proteins and ligands and establish more precise methods to determine the dissociation constant *"in crystallo"*.

6.4 Conclusion

X-ray crystallography has been shown to be a useful and powerful technique in molecular biology and biochemistry to understand the interaction between a protein

and its ligands, receptors or inhibitors, at the atomic level. In this thesis, the structures and ligand binding interactions of BLG and CYP3 have been studied.
The binding of BLG with a variety of hydrophobic ligands has been studied in solution for many years. However, the actual binding site of these small molecules, such as fatty acid, are arguable as there are two possible binding site to accommodate a ligand; one is in the calyx inside the β-barrel and the other is in the groove on the protein surface. The work presented here is the first direct observation of a ligand binding to BLG. A detailed description of the interaction between a fatty acid (palmitate) and BLG can therefore be examined. Moreover, the structure of palmitate-bound BLG could provide an insight into the biological phenomenon observed by other methods, such as fluorescence and site-mutagenesis study.
As for CYP3, this is the first structure of a cyclophilin from the free-living nematode *C. elegans* to be solved. The structure of CYP3 has provided many informations to understand the various biological roles that cyclophilin plays and may help the development of targeted antiparasite drugs, for example, CsA analogues.
Furthermore, the structure of CYP3 complexed with Ala-Pro has been shown to be a good model system to study the ligand binding behaviour *"in crystallo"*. This is the first report to obtain the dissociation constant *"in crystallo"* by x-ray crystallography. This result has been shown to be in agreement with the result obtained from the PPIase assay in solution. Therefore, this experimental system may provide a starting point to develop a method to determine a dissociation constant by x-ray crystallography when other techniques, for example, fluorescence, are insensitive or unavailable.

7. Bibliography and Appendix

Ali, S. and Clark, A. J. (1988) *Journal Of Molecular Biology* **199**, 415-426

Anderson, S. K., Gallinger, S., Roder, J., Frey, J., Young, H. A., and Ortaldo, J. R. (1993) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **90**, 542-546

Armbruster, A. M. and Pullman, A. (1974) *Febs Letters* **49**, 18-21

Armstrong, J. M. and McKenzie, H. A. (1967) *Biochimica Et Biophysica Acta* **147**, 60-72

Aschaffenberg, R. and Drewry, J. (1955) *Nature* **176**, 218-219

Aschaffenburg, R., Green, D. W., and Simmons, R. M. (1965) *Journal Of Molecular Biology* **13**, 194-201

Bachinger, H. P. (1987) *Journal Of Biological Chemistry* **262**, 17144-17148

Bell, A., Roberts, H. C., and Chappell, L. H. (1996) *General Pharmacology* **27**, 963-971

Bell, A., Wernli, B., and Franklin, R. M. (1994) *Biochemical Pharmacology* **48**, 495-503

Bell, K. and McKenzie, H. A. (1967) *Biochimica Et Biophysica Acta* **147**, 109-122

Bell, K., McKenzie, H. A., Muller, V., Rogers, C., and Shaw, D. C. (1981) *Comparative Biochemistry And Physiology B-Biochemistry & Molecular Biology* **68**, 225-236

Bell, K., McKenzie, H. A., Murphy, W. H., and Shaw, D. C. (1970) *Biochim. Biophys. Acta* **214**, 427-436

Bell, K., McKenzie, H. A., and Shaw, D. C. (1981) *Australian Journal Of Biological Sciences* **34**, 133-147

Bergsma, D. J., Eder, C., Gross, M., Kersten, H., Sylvester, D., Appelbaum, E., Cusimano, D., Livi, G. P., Mclaughlin, M. M., Kasyan, K., Porter, T. G., Silverman, C., Dunnington, D., Hand, A., Prichett, W. P., Bossard, M. J., Brandt, M., and Levy, M. A. (1991) *Journal Of Biological Chemistry* **266**, 23204-23214

Beveridge, T. and Calne, R. Y. (1995) *Transplantation* **59**, 1568-1570

Bewley, M. C., Qin, B. Y., Jameson, G. B., Sawyer, L., and Baker, E. N. (1997) in *Milk protein polymorphism* pp.100-109, International Dairy Federation, Belgium

Blundell, T. L. and Johnson, L. N. (1976) in *Protein crystallography* (Horecker, B., Kaplan, N. O., Marmur, J., and Scheraga, H. eds) pp. 59-68, Academic Press, London

Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E. V., Cavaggioni, A., Findlay, J. B. C., and North, A. C. T. (1992) *Nature* **360**, 186-188

Bolognesi, M., Liberatori, J., Oberti, R., and Ungaretti, L. (1979) *Journal Of Molecular Biology* **131**, 411-413

Braaten, D., Franke, E. K., and Luban, J. (1996) *Journal Of Virology* **70**, 3551-3560

Brandts, J. F. (1975) *Biochemistry* **14**, 4953-4963

Briggs, D. R. and Hull, R. (1945) *Journal Of The American Chemical Society* **67**, 2007-2011

Brignon, G., Ridadeau-Dumas, B., Garnier, J., Pantaloni, D., Guinard, S., Basch, J. J., and Timasheff, S. N. (1969) *Archives Of Biochemistry And Biophysics* **129**, 720-727

Brownlow, S., Cabral, J. H. M., Cooper, R., Flower, D. R., Yewdall, S. J., Polikarpov, I., North, A. C. T., and Sawyer, L. (1997) *Structure* **5**, 481-495

Brunger, A. T. (1990a) *Acta Crystallographica Section A* **46**, 46-57

Brunger, A. T., Krukowski, A., and Erickson, J. W. (1990b) *Acta Crystallographica Section A* **46**, 585-593

Brunger, A. T. (1992) in *X-PLOR. A system for X-ray crystallography and NMR.* (AnonymousYale University Press, New Haven, Connecticut.

Brunger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science* **235**, 458-460

Bull, H. B. (1946) *Journal Of The American Chemical Society* **68**, 747-748

Cabral, M. (1993) *PhD Thesis, University of Edinburgh*

Chen, R. F. (1967) *Journal Of Biological Chemistry* **242**, 173-177

Chen, S. X., Hardin, C. C., and Swaisgood, H. E. (1993) *Journal Of Protein Chemistry* **12**, 613-625

Cho, Y. J., Batt, C. A., and Sawyer, L. (1994) *Journal Of Biological Chemistry* **269**, 11102-11107

Clipstone, N. A., Fiorentino, D. F., and Crabtree, G. R. (1994) *Journal Of Biological Chemistry* **269**, 26431-26437

Colgan, J., Yuan, H. E. H., Franke, E. K., and Luban, J. (1996) *Journal Of Virology* **70**, 4299-4310

Colley, N. J., Baker, E. K., Stamnes, M. A., and Zuker, C. S. (1991) *Cell* **67**, 255-263

Cowan, S. W., Newcomer, M. E., and Jones, T. A. (1990) *Proteins-Structure Function And Genetics* **8**, 44-61

Creamer, L. K. (1995) *Biochemistry* **34**, 7170-7176

Crowfoot, D. M. and Riley, D. P. (1938) *Nature* **141**, 521-522

Crowther, R. A. and Blow, D. M. (1967) *Acta Crystallographica* **23**, 544-548

Dauter, Z. (1997) *Methods In Enzymology* **276**, 326-344

Davis, J. M., Boswell, B. A., and Bachinger, H. P. (1989) *Journal Of Biological Chemistry* **264**, 8956-8962

Dewan, J. C. and Tilton, R. F. (1987) *Journal Of Applied Crystallography* **20**, 130-132

Dodson, E. (1997) *CCP4 program manual*

Dornan, J. (1999a) *Personal communication*

Dornan, J., Page, A. P., Taylor, P., Wu, S. Y., Winter, A. D., Husi, H., and Walkinshaw, M. D. (1999b) *Journal Of Biological Chemistry* (in press)

Dufour, E. and Haertle, T. (1990a) *Protein Engineering* **4**, 185-190

Dufour, E. and Haertle, T. (1990b) *Journal Of Agricultural And Food Chemistry* **38**, 1691-1695

Dufour, E. and Haertle, T. (1991) *Biochimica Et Biophysica Acta* **1079**, 316-320

Dunnill, P. and Green, D. W. (1965) *Journal Of Molecular Biology* **15**, 147-151

Eberhardt, E. S., Loh, S. N., Hinck, A. P., and Raines, R. T. (1992) *Journal Of The American Chemical Society* **114**, 5437-5439

Erhardt, G., Godovaczimmermann, J., and Conti, A. (1989) *Biological Chemistry Hoppe-Seyler* **370**, 757-762

Ferreira, P. A., Nakayama, T. A., Pak, W. L., and Travis, G. H. (1996) *Nature* **383**, 637-640

Fischer, G., Bang, H., and Mech, C. (1984) *Biomedica Biochimica Acta* **43**, 1101-1111

Fischer, G., Berger, E., and Bang, H. (1989a) *Febs Letters* **250**, 267-270

Fischer, G., Wittmannliebold, B., Lang, K., Kiefhaber, T., and Schmid, F. X. (1989b) *Nature* **337**, 476-478

Fischer, S., Dunbrack, R. L., and Karplus, M. (1994) *Journal Of The American Chemical Society* **116**, 11931-11937

Flower, D. R. (1996) *Biochemical Journal* **318**, 1-14

Fogolari, F., Ragiba, L., Romagnoli, S., Druif, K. D., and Molinari, H. (1998) *Febs Letters* **436**, 149-154

Franke, E. K., Yuan, H. E. H., and Luban, J. (1994) *Nature* **372**, 359-362

Fransson, C., Freskgard, P. O., Herbertsson, H., Johansson, A., Jonasson, P., Martensson, L. G., Svensson, M., Jonsson, B. H., and Carlsson, U. (1992) *Febs Letters* **296**, 90-94

Frapin, D., Dufour, E., and Haertle, T. (1993) *Journal Of Protein Chemistry* **12**, 443-449

Fugate, R. D. and Song, P. S. (1980) *Biochimica Et Biophysica Acta* **625**, 28-42

Futterman, S. and Heller, J. (1972) *Journal Of Biological Chemistry* **247**, 5168-5172

Galat, A. and Metcalfe, J. (1995) *Progress in biophysics and molecular biology* **63**, 67-118

Gamble, T. R., Vajdos, F. F., Yoo, S. H., Worthylake, D. K., Houseweart, M., Sundquist, W. I., and Hill, C. P. (1996) *Cell* **87**, 1285-1294

Garman, E. F. (1996) in *Crystallographic methods and protocols* (Jones, C., Mulloy, B., and Sanderson, M. R. eds) pp. 119-123, Humana Press, Totowa, New Jersey

Geroges, C. and Guinand, S. (1960) *J. Chem. Phys.* **57**, 606-614

Glusker, J. P., Lewis, M., and Rossi, M. (1994a) in *Crystal structure analysis for chemists and biologists* pp. 389-411, VCH publishers, New York

Glusker, J. P., Lewis, M., and Rossi, M. (1994b) in *Crystal structure analysis for chemists and biologists* pp. 345-388, VCH publishers, New York

Godovaczimmermann, J., Krause, I., Buchberger, J., Weiss, G., and Klostermeyer, H. (1990) *Biological Chemistry Hoppe-Seyler* **371**, 255-260

Green, D. W., Aschaffenberg, R., Gamerman, A., Coppola, J. C., Dunnill, P., Simmons, R. M., Komorowski, E. S., Sawyer, L., Turner, E. M. C., and Woods, K. F. (1979) *Journal Of Molecular Biology* **131**, 375-397

Green, D. W., North, A. C. T., and Aschaffenberg, R. (1956) *Biochimica Et Biophysica Acta* **21**, 583-585

Grosclaude, F., Mahe, M. F., Mercier, J. C., Bonnemaire, J., and Tessier, J. H. (1976) *Ann. Genet. Sel. Anim.* **8**, 461-476

Gumpen, S., Hegg, P. O., and Martens, H. (1979) *Biochimica Et Biophysica Acta* **574**, 189

Halliday, J. A., Bell, K., McAndrew, K., and Shaw, D. C. (1993) *Protein sequences and data analysis* **5**, 201-205

Hambling, S. G. (1990) *PhD Thesis, University of Edinburgh*

Hambling, S. G., McAlpine, A. S., and Sawyer, L. (1992) *in Advanced Dairy Chemistry, Vol. 1 (Fox, P. F. ed), Elservier Applied Science, London* 140-190

Handschumacher, R. E., Harding, M. W., Rice, J., and Drugge, R. J. (1984) *Science* **226**, 544-547

Harding, M. W., Handschumacher, R. E., and Speicher, D. W. (1986) *Journal Of Biological Chemistry* **261**, 8547-8555

Harrison, R. K. and Stein, R. L. (1990a) *Biochemistry* **29**, 1684-1689

Harrison, R. K. and Stein, R. L. (1990b) *Biochemistry* **29**, 3813-3816

Hasel, K. W., Glass, J. R., Godbout, M., and Sutcliffe, J. G. (1991) *Molecular And Cellular Biology* **11**, 3484-3491

Hendrickson, W. A. (1985) *Methods In Enzymology* **115**, 252-270

Hinrichs, D. J., Wegmann, K. W., and Peters, B. A. (1983) *Cellular Immunology* **77**, 202-209

Hodsdon, M. E. and Cistola, D. P. (1997a) *Biochemistry* **36**, 1450-1460

Hodsdon, M. E. and Cistola, D. P. (1997b) *Biochemistry* **36**, 2278-2290

Hoffmann, K., Armitage, I. M., and Handschumacher, R. E. (1995a) *Faseb Journal* **9**, A1289

Hoffmann, K. and Handschumacher, R. E. (1995b) *Biochemical Journal* **307**, 5-8

Holden, H. M., Rypniewski, W. R., Law, J. H., and Rayment, I. (1987) *Embo Journal* **6**, 1565-1570

Huber, R., Schneider, M., Mayr, I., Muller, R., Deutzmann, R., Suter, F., Zuber, H., Falk, H., and Kayser, H. (1987) *Journal Of Molecular Biology* **198**, 499-513

Jack, A. and Levitt, M. (1978) *Acta Crystallographica Section A* **A34**, 931-935

Jackson, S. E. and Fersht, A. R. (1991) *Biochemistry* **30**, 10436-10443

Jancarik, J. and Kim, S. H. (1991) *Journal Of Applied Crystallography* **24**, 409

Jang, H. D. and Swaisgood, H. E. (1990) *Journal Of Dairy Science* **73**, 2067-2074

Johnson, J. L. and Toft, D. O. (1994) *Journal Of Biological Chemistry* **269**, 24989-24993

Jones, T. A. (1992) *Proceedings of the Daresbury study weekend on Moleculare replacement* 91-105

Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallographica Section A-Fundamentals Of Crystallography* **47**, 110-119

Jorgensen, W. L. and Gao, J. (1988) *Journal Of The American Chemical Society* **110**, 4212-4216

Kabsch, W. and Sander, C. (1983) *Biopolymers* **22**, 2577-2637

Kallen, J., Mikol, V., Taylor, P., and Walkinshaw, M. D. (1998) *Journal Of Molecular Biology* **283**, 435-449

Kallen, J., Spitzfaden, C., Zurini, M. G. M., Wider, G., Widmer, H., Wuthrich, K., and Walkinshaw, M. D. (1991) *Nature* **353**, 276-279

Kallen, J. and Walkinshaw, M. D. (1992) *Febs Letters* **300**, 286-290

Katakura, Y., Totsuka, M., Ametani, A., and Kaminogawa, S. (1994) *Biochimica Et Biophysica Acta-Protein Structure And Molecular Enzymology* **1207**, 58-67

Ke, H. M. (1992) *Journal Of Molecular Biology* **228**, 539-550

Ke, H. M., Mayrose, D., Belshaw, P. J., Alberg, D. G., Schreiber, S. L., Chang, Z. Y., Etzkorn, F. A., Ho, S., and Walsh, C. T. (1994) *Structure* **2**, 33-44

Ke, H. M., Mayrose, D., and Cao, W. (1993a) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **90**, 3324-3328

Ke, H. M., Zhao, Y. D., Luo, F., Weissman, I., and Friedman, J. (1993b) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **90**, 11850-11854

Ke, H. M., Zydowsky, L. D., Liu, J., and Walsh, C. T. (1991) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **88**, 9483-9487

Kelywegt, G. J. and Jones, A. T. (1996) *Rave-the manual*

Kieffer, L. J., Thalhammer, T., and Handschumacher, R. E. (1992) *Journal Of Biological Chemistry* **267**, 5503-5507

Kiefhaber, T., Grunert, H. P., Hahn, U., and Schmid, F. X. (1990a) *Biochemistry* **29**, 6475-6480

Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F. X. (1990b) *Biochemistry* **29**, 3061-3070

Kiefhaber, T., Quaas, R., Hahn, U., and Schmid, F. X. (1990c) *Biochemistry* **29**, 3053-3061

Kleywegt, G. J. (1996) *ESF/CCP4 Newsletter* **32**, 32-36

Kofron, J. L., Kuzmic, P., Kishore, V., Colonbonilla, E., and Rich, D. H. (1991) *Biochemistry* **30**, 6127-6134

Kolde, H. J. and Braunitzer, G. (1983) *Milchwissenschaft-Milk Science International* **38**, 70-72

Konno, M., Ito, M., Hayano, T., and Takahashi, N. (1996) *Journal Of Molecular Biology* **256**, 897-908

Kontopidis, G. (1999) *Personal communication*

Kraulis, P. J. (1991) *Journal Of Applied Crystallography* **24**, 946-950

Krause, I., Buchberger, J., Weiss, G., Pflugler, M., and Klostermeyer, H. (1988) *Electrophoresis* **9**, 609-613

Lang, K., Schmid, F. X., and Fischer, G. (1987) *Nature* **329**, 268-270

Lange, D. C., Kothari, R., Patel, R. C., and Patel, S. C. (1998) *Biophysical Chemistry* **74**, 45-51

Lapanje, S. and Poklar, N. (1989) *Biophysical Chemistry* **34**, 155

Larson, B. L. (1979) *Journal Of Dairy Research* **46**, 161-74

Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *Journal Of Applied Crystallography* **26**, 283-291

Liu, J., Albers, M. W., Chen, C. M., Schreiber, S. L., and Walsh, C. T. (1990) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **87**, 2304-2308

Liu, J., Albers, M. W., Wandless, T. J., Luan, S., Alberg, D. G., Belshaw, P. J., Cohen, P., Mackintosh, C., Klee, C. B., and Schreiber, S. L. (1992) *Biochemistry* **31**, 3896-3901

Liu, J., Farmer, J. D., Lane, W. S., Friedman, J., Weissman, I., and Schreiber, S. L. (1991) *Cell* **66**, 807-815

Loh, C., Shaw, K. T. Y., Carew, J., Viola, J. P. B., Luo, C., Perrino, B. A., and Rao, A. (1996) *Journal Of Biological Chemistry* **271**, 10884-10891

Loosli, H. R., Kessler, H., Oschkinat, H., Weber, H. P., Petcher, T. J., and Widmer, A. (1985) *Helvetica Chimica Acta* **68**, 682-704

Luo, C., Shaw, K. T. Y., Raghavan, A., Aramburu, J., Garciacozar, F., Perrino, B. A., Hogan, P. G., and Rao, A. (1996) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **93**, 8907-8912

Lyster, R. L. J. (1972) *Journal Of Dairy Research* **39**, 279-318

Marden, M. C., Dufour, E., Christova, P., Huange, Y., Hostis, E. L., and Haertle, T. (1994) *Archives Of Biochemistry And Biophysics* **311**, 258-262

McAlpine, A. S. (1991) *PhD Thesis, University of Edinburgh*

McDougall, E. I. and Stewart, J. C. (1976) *Biochemical Journal* **153**, 647-655

McKenzie, H. A. (1971) *Milk Proteins-II, Academic Press, New York* 257-330

McKenzie, H. A., Muller, V. J., and Treacy, G. B. (1983) *Comparative Biochemistry And Physiology B-Biochemistry & Molecular Biology* **74**, 259-271

McKenzie, H. A., Sawyer, W. H., and Smith, M. B. (1967) *Biochimica Et Biophysica Acta* **147**, 73-92

McRee, D. E. (1993) in *Practical protein crystallography* pp. 172-176, Academic press, San Diego

Mikol, V., Kallen, J., Pflugl, G., and Walkinshaw, M. D. (1993) *Journal Of Molecular Biology* **234**, 1119-1130

Mikol, V., Kallen, J., and Walkinshaw, M. D. (1994) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **91**, 5183-5186

Monaco, H. L., Zanotti, G., Coda, A., Bignetti, E., Ramoni, R., Grolli, S., Bedarker, S., Cachau, R., and Amzel, L. M. (1990) *Suppl. Minerva Biotecnologica* **2**, 150-153

Monaco, H. L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L., and Eliopoulos, E. E. (1987) *Journal Of Molecular Biology* **197**, 695-706

Narayan, M. and Berliner, L. J. (1997) *Biochemistry* **36**, 1906-1911

Navaza, J. and Saludjian, P. (1997) *Methods In Enzymology* **276**, 581-594

Newcomer, M. E. (1993) *Structure* **1**, 7-18

Newcomer, M. E., Jones, T. A., Aqvist, J., Sundelin, J., Eriksson, U., Rask, L., and Peterson, P. A. (1984) *Embo Journal* **3**, 1451-1454

Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins-Structure Function And Genetics* **11**, 281-296

North, A. C. T. (1989) *International Journal Of Biological Macromolecules* **11**, 56-58

Nussenblatt, R. B., Salinascarmona, M., Waksman, B. H., and Gery, I. (1983) *International Archives Of Allergy And Applied Immunology* **70**, 289-294

Oneill, T. E. and Kinsella, J. E. (1987) *Journal Of Agricultural And Food Chemistry* **35**, 770-774

Osterhoff, D. R. and Pretorius, A. M. G. (1966) *Proc. S. Afri. Soc. Anim. Prod* **5**, 531-536

Otwinowski, Z. and Minor, W. (1997) *Methods In Enzymology* **276**, 307-326

Page, A. P., Kumar, S., and Carlow, C. K. S. (1995a) *Parasitology Today* **11**, 385-388

Page, A. P., Landry, D., Wilson, G. G., and Carlow, C. K. S. (1995b) *Biochemistry* **34**, 11545-11550

Page, A. P., MacNiven, K., and Hengartner, M. O. (1996) *Biochemical Journal* **317**, 179-185

Page, A. P. and Winter, A. D. (1998) *Molecular And Biochemical Parasitology* **95**, 215-227

Palmer, A. H. (1934) *Journal Of Biological Chemistry* **104**, 359-372

Papiz, M. Z. (1982) *PhD Thesis, Napier University*

Papiz, M. Z., Sawyer, L., Eliopoulos, E. E., North, A. C. T., Findlay, J. B. C., Sivaprasadarao, R., Jones, T. A., Newcomer, M. E., and Kraulis, P. J. (1986) *Nature* **324**, 383-385

Park, K. H. and Lund, D. B. (1984) *Journal Of Dairy Science* **67**, 1699-1706

Pedersen, K. O. (1936) *Biochemical Journal* **30**, 948-960

Perez, M. D. and Calvo, M. (1995) *Journal Of Dairy Science* **78**, 978-988

Perez, M. D., Devillegas, C. D., Sanchez, L., Aranda, P., Ena, J. M., and Calvo, M. (1989) *Journal Of Biochemistry* **106**, 1094-1097

Perez, M. D., Puyol, P., Ena, J. M., and Calvo, M. (1993) *Journal Of Dairy Research* **60**, 55-63

Perez, M. D., Sanchez, L., Aranda, P., Ena, J. M., Oria, R., and Calvo, M. (1991) *Biochimica Et Biophysica Acta* **1123**, 151-155

Pervais, S. and Brew, K. (1987) *Faseb Journal* **1**, 209-214

Pervaiz, S. and Brew, K. (1986) *Archives Of Biochemistry And Biophysics* **246**, 846-854

Pflugl, G., Kallen, J., Schirmer, T., Jansonius, J. N., Zurini, M. G. M., and Walkinshaw, M. D. (1993) *Nature* **361**, 91-94

Polikarpov, I. (1995) *Personal communication*

Price, E. R., Zydowsky, L. D., Jin, M. J., Baker, C. H., Mckeon, F. D., and Walsh, C. T. (1991) *Proceedings Of The National Academy Of Sciences Of The United States Of America* **88**, 1903-1907

Puyol, P., Perez, M. D., Ena, J. M., and Calvo, M. (1991) *Agricultural And Biological Chemistry* **55**, 2515-2520

Puyol, P., Perez, M. D., Peiro, J. M., and Calvo, M. (1994) *Journal Of Dairy Science* **77**, 1494-1502

Qi, X. L., Brownlow, S., Holt, C., and Sellers, P. (1995) *Biochimica Et Biophysica Acta-Protein Structure And Molecular Enzymology* **1248**, 43-49

Qin, B. Y., Bewley, M. C., Creamer, L. K., Baker, E. N., and Jameson, G. B. (1999) *Protein Science* **8**, 75-83

Qin, B. Y., Bewley, M. C., Creamer, L. K., Baker, H. M., Baker, E. N., and Jameson, G. B. (1998a) *Biochemistry* **37**, 14014-14023

Qin, B. Y., Creamer, L. K., Baker, E. N., and Jameson, G. B. (1998b) *Febs Letters* **438**, 272-278

Radzicka, A., Acheson, S. A., and Wolfenden, R. (1992) *Bioorganic chemistry* **20**, 382-386

Ragona, L., Pusterla, F., Zetta, L., Monaco, H. L., and Molinari, H. (1997) *Folding & Design* **2**, 281-290

Relkin, P., Launay, B., and Eynard, L. (1993) *Journal Of Dairy Science* **76**, 36

Ries-Kautt, M. and Ducruix, A. (1992) in *Crystallization of nucleic acids and proteins* (Ducruix, A. and Giege, R. eds) pp. 195-218, IRL Press, Oxford

Rose, G. D., Gierasch, L. M., and Smith, J. A. (1985) *Advances in protein chemistry* **37**, 1-109

Rosenwirth, B., Billich, A., Datema, R., Donatsch, P., Hammerschmid, F., Harrison, R., Hiestand, P., Jaksche, H., Mayer, P., Peichl, P., Quesniaux, V., Schatz, F., Schuurman, H. J., Traber, R., Wenger, R., Wolff, B., Zenke, G., and Zurini, M. (1994) *Antimicrobial Agents And Chemotherapy* **38**, 1763-1772

Rossmann, M. G. and Blow, D. M. (1962) *Acta Crystallographica Section B* **15**, 24-31

Sacchettini, J. C., Scapin, G., Gopaul, D., and Gordon, J. I. (1992) *Journal Of Biological Chemistry* **267**, 23534-23545

Sawyer, L., Brownlow, S., Polikarpov, I., and Wu, S. Y. (1998) *International Dairy Journal* **8**, 65-72

Sawyer, W. H., Norton, R. S., Nichol, L. W., and McKenzie, G. H. (1971) *Biochim. Biophys. Acta* **243**, 19

Schmidt, B., Rahfeld, J., Schierhorn, A., Ludwig, B., Hacker, J., and Fischer, G. (1994) *Febs Letters* **352**, 185-190

Schonbrunner, E. R., Mayer, S., Tropschug, M., Fischer, G., Takahashi, N., and Schmid, F. X. (1991) *Journal Of Biological Chemistry* **266**, 3630-3635

Senti, F. R. and Warner, R. C. (1948) *Journal Of The American Chemical Society* **70**, 3318-3320

Sheldrick, G. M. and Schneider, T. R. (1997) *Methods In Enzymology* **277**, 319-343

Smith, T., Ferreira, L. R., Hebert, C., Norris, K., and Sauk, J. J. (1995) *Journal Of Biological Chemistry* **270**, 18323-18328

Spector, A. A. and Fletcher, J. E. (1970) *Lipids* **5**, 403-411

Spitzfaden, C., Weber, H. P., Braun, W., Kallen, J., Wider, G., Widmer, H., Walkinshaw, M. D., and Wuthrich, K. (1992) *Febs Letters* **300**, 291-300

Steinmann, B., Bruckner, P., and Supertifurga, A. (1991) *Journal Of Biological Chemistry* **266**, 1299-1303

Stiller, C. R., Dupre, J., Gent, M., Jenner, M. R., Keown, P. A., Laupacis, A., Martell, R., Rodger, N. W., Vongraffenried, B., and Wolfe, B. M. J. (1984) *Science* **223**, 1362-1367

Stura, E. A. and Chen, P. (1992) in *Crystallization of Nucleic acids and proteins; A Practical Approach* (Ducruix, A. and Giege, R. eds) pp. 241-254, IRL Press, Oxford Univeristy

Stura, E. A. and Wilson, I. A. (1992) in *Crystallization of Nucleic Acids and Proteins; A Practical Approach* (Ducruix, A. and Giege, R. eds) pp. 100-101, IRL Press, Oxford Univeristy

Tanford, C., Bunville, L. G., and Nozaki, Y. (1959) *Journal Of The American Chemical Society* **81**, 4032-4036

Taylor, P., Husi, H., Kontopidis, G., and Walkinshaw, M. D. (1997) *Progress in biophysics and molecular biology* **67**, 155-181

Taylor, P. (1998a) *Personal communication*

Taylor, P., Page, A. P., Kontopidis, G., Husi, H., and Walkinshaw, M. D. (1998b) *Febs Letters* **425**, 361-366

Thalhammer, T., Kieffer, L. J., Jiang, T. R., and Handschumacher, R. E. (1992) *European Journal Of Biochemistry* **206**, 31-37

Thali, M., Bukovsky, A., Kondo, E., Rosenwirth, B., Walsh, C. T., Sodroski, J., and Gottlinger, H. G. (1994) *Nature* **372**, 363-365

Theriault, Y., Logan, T. M., Meadows, R., Yu, L. P., Olejniczak, E. T., Holzman, T. F., Simmer, R. L., and Fesik, S. W. (1993) *Nature* **361**, 88-91

Tickle, I. J. and Driessen, H. P. C. (1996) in *Crystallographic methods and protocols* (Jones, C., Mulloy, B., and Sanderson, M. R. eds) pp. 173-203, Humana Press, Totowa, New Jersey

Tilley, J. M. A. (1960) *Dairy science abstract* **22**, 111-125

Timasheff, S. N., Townend, R., and Mescanti, L. (1966) *Journal Of Biological Chemistry* **241**, 1863-1870

Townend, R. and Basch, J. J. (1968) *Archives Of Biochemistry And Biophysics* **126**, 59

Townend, R., Herskovits, T. T., Timasheff, S. N., and Gorbunoff, M. J. (1969) *Archives Of Biochemistry And Biophysics* **129**, 567-580

Townend, R., Weinberger, L., and Timasheff, S. N. (1960) *Journal Of The American Chemical Society* **82**, 3168-3174

Weber, C., Wider, G., Vonfreyberg, B., Traber, R., Braun, W., Widmer, H., and Wuthrich, K. (1991) *Biochemistry* **30**, 6563-6574

Weber, P. C. (1997) *Methods In Enzymology* **276**, 13-22

Westhof, E. and Dumas, P. (1996) in *Crystallographic methods and protocols* (Jones, C., Mulloy, B., and Sanderson, M. R. eds) Humana Press, Totowa, New Jersey

Wilmot, C. M. and Thornton, J. M. (1988) *Journal Of Molecular Biology* **203**, 221-232

Wuthrich, K., Spitzfaden, C., Memmert, K., Widmer, H., and Wider, G. (1991) *Febs Letters* **285**, 237-247

Yeh, J. I. and Hol, W. G. J. (1998) *Acta Crystallographica Section D-Biological Crystallography* **54**, 479-480

Yewdall, S. J. (1988) *PhD Thesis, University of Leeds*

Zhao, Y. D., Chen, Y. Q., Schutkowski, M., Fischer, G., and Ke, H. M. (1997) *Structure* **5**, 139-146

Zhao, Y. D. and Ke, H. M. (1996a) *Biochemistry* **35**, 7356-7361

Zhao, Y. D. and Ke, H. M. (1996b) *Biochemistry* **35**, 7362-7368

APPENDIX

Parts of the works described in this thesis have been published in the following articles:

Dornan, J., Page, A. P., Taylor, P., Wu, S. Y., Winter, A. D., Husi, H., and Walkinshaw, M. D. (1999) *Journal Of Biological Chemistry* (in press)

Sawyer, L., Brownlow, S., Polikarpov, I., and Wu, S. Y. (1998) *International Dairy Journal* **8**, 65-72

Sawyer, L., Kontopidis, G., and Wu, S. Y. (1999) *International Journal Of Food Science and Technology* **34**, 409-418

Wu, S. Y., Perez, M. D., Puyol, P., and Sawyer, L. (1999) *Journal Of Biological Chemistry* **274**, 170-174

Appendix 2-1

(1) Systematic absence of BLGY data (C2221)
(absence)

Intensities of systematic absences

h	k	l	Intensity	Sigma	I/Sigma
0	0	3	627.3	113.6	5.5
0	0	5	52.7	130.0	0.4
0	0	7	625.1	161.8	3.9
0	0	9	119.2	114.9	1.0
0	0	11	65.2	156.4	0.4
0	0	13	-155.7	209.5	-0.7
0	0	17	-72.0	185.6	-0.4
0	0	19	623.2	270.8	2.3
0	0	21	33.9	195.9	0.2
0	0	23	463.5	219.2	2.1
0	0	25	276.5	423.4	0.7
0	0	27	474.1	217.6	2.2
0	0	29	452.4	427.3	1.1
0	0	31	-32.9	243.4	-0.1
0	0	33	211.8	450.3	0.5

(present)

0	0	4	250472	29790
0	0	6	2607.9	175.2
0	0	8	12540.7	742.8
0	0	10	10936.7	553.6
0	0	12	32877.1	1383.4
0	0	14	127262	4144
0	0	16	285866	13724
0	0	18	82921.8	4774.4
0	0	20	396.8	243.2
0	0	22	4181.9	335.8
0	0	24	818.2	311.0
0	0	26	9971.3	1146.2
0	0	28	2646.0	377.6
0	0	30	1046.7	281.5
0	0	32	21811.2	5512.2

====== > reflection condition (0 0 2n)

(2) Systematice absence of BLG co-crystallized with Palmitic acid (P3221)

(absence)

Intensities of systematic absences

h	k	l	Intensity	Sigma	I/Sigma
0	0	13	-58.2	35.6	-1.6
0	0	14	-1.9	38.7	0.0
0	0	16	90.3	44.5	2.0
0	0	17	-25.4	33.3	-0.8
0	0	19	-74.2	57.1	-1.3
0	0	20	-35.1	62.9	-0.6
0	0	22	60.8	92.3	0.7
0	0	25	121.4	77.3	1.6
0	0	26	74.9	75.9	1.0
0	0	28	155.8	77.2	2.0
0	0	29	53.3	61.2	0.9
0	0	31	-87.7	57.8	-1.5
0	0	32	106.3	82.1	1.3
0	0	34	-18.3	82.1	-0.2
0	0	35	163.1	87.8	1.9
0	0	37	12.9	90.5	0.1
0	0	38	6.1	69.2	0.1
0	0	40	79.9	97.9	0.8
0	0	41	69.3	90.9	0.8
0	0	43	176.0	106.5	1.7
0	0	44	-154.4	113.8	-1.4

(present)

0	0	12	5896.1	392.5
0	0	15	16379.6	1030.2
0	0	18	3945.4	353.6
0	0	21	14225.1	1252.5
0	0	27	903.0	137.2
0	0	30	4348.2	533.7
0	0	33	1147.2	182.8
0	0	36	709.0	147.5
0	0	39	2920.4	698.3
0	0	42	-48.0	87.3

====== > reflection condition (0 0 3n)

===> space group : p3121 or p3221

(1) Systematic absence of CYP3 data (P41212)
systematic absence:
(a) absence:
Intensities of systematic absences

h	k	l	Intensity	Sigma	I/Sigma
0	0	21	-22.4	44.6	-0.5
0	0	22	-64.9	49.1	-1.3
0	0	23	-5.9	34.8	-0.2
0	0	26	23.4	40.0	0.6
0	0	27	169.6	61.4	2.8
0	0	29	34.6	46.5	0.7
0	0	31	-34.4	57.0	-0.6
0	0	33	171.4	84.2	2.0
0	0	34	40.8	57.3	0.7
0	0	35	164.3	85.2	1.9
0	0	37	0.6	60.4	0.0
0	0	38	18.8	84.1	0.2
0	0	39	-60.1	84.2	-0.7
0	0	41	-153.2	84.6	-1.8
0	0	42	-145.3	59.5	-2.4
0	0	43	-177.8	82.9	-2.1
0	0	45	-65.7	71.3	-0.9
0	0	46	27.9	80.0	0.3
0	0	47	2.2	81.0	0.0
0	0	49	48.7	81.0	0.6
0	0	50	93.2	59.1	1.6
0	0	51	7.5	78.9	0.1
0	0	53	-15.0	55.0	-0.3
0	0	54	54.0	82.3	0.7
0	0	55	-28.0	81.7	-0.3
0	0	57	142.8	90.1	1.6
0	0	58	75.4	59.8	1.3
0	0	59	217.1	102.1	2.1
0	0	61	-145.3	78.3	-1.9
0	0	62	-55.8	79.5	-0.7
0	0	63	45.4	56.6	0.8
0	0	65	63.9	77.5	0.8
0	0	66	39.8	52.1	0.8
0	0	67	-5.4	79.8	-0.1
5	0	0	7.0	12.5	0.6
7	0	0	12.6	16.8	0.7
9	0	0	-19.2	22.3	-0.9
11	0	0	33.9	25.9	1.3
13	0	0	-19.4	44.6	-0.4
17	0	0	145.6	69.2	2.1
19	0	0	-98.6	64.8	-1.5
21	0	0	-0.8	60.8	0.0
23	0	0	-65.5	59.5	-1.1
27	0	0	-44.6	62.2	-0.7

(b) present:

0	0	20	7457.2	623.9
0	0	24	43631.1	3479.1
0	0	28	11081.0	1040.7
0	0	32	5932.4	709.7
0	0	36	4962.0	580.8
0	0	40	33515.2	4328.8
0	0	44	9621.1	1635.7
0	0	48	46902.0	8657.8
0	0	52	32118.0	6999.6
0	0	56	2131.5	566.7
0	0	60	311.6	139.8
0	0	64	183.3	108.0
0	0	68	1563.2	868.4
6	0	0	16036.1	993.3
8	0	0	677.0	47.7
10	0	0	605.6	44.1
12	0	0	9846.0	788.1
14	0	0	324.4	62.4
16	0	0	1167.6	151.9
18	0	0	3221.9	431.9
20	0	0	2261.3	343.2
22	0	0	2051.4	355.9
24	0	0	2221.7	446.5

===> relection condition : (0 0 4n) & (2n 0 0)
===> space group : p41212 or p43212